

Bologna, Britain and academic freedom

The British government is in difficulties in devising a form of words for the definition of academic freedom, which is part of the cost of seeking too much central control.

THE University of Bologna, this year celebrating its ninth centenary, is the prototype of the European university and should therefore have been a greater inspiration to the British government, and its Secretary of State for Education and Science, Mr Kenneth Baker, in the preparation of the Education Reform Bill, one crucial clause of which was defeated in the House of Lords last week. For Bologna's origins would ordinarily have commanded the particular respect of the present British government — the wish of the Italian meritocracy of the time to give its sons a good start in life. The legend is that successful families provided their young with the funds to hire a bunch of scholars to instruct them in the secular ingredients of higher learning. The arrangement made evident economic sense for the customers. The first modern university was a thorough model of the market economy in higher education; several decades went by before teachers usurped students from control of Bologna's governing councils.

If the British government had modelled the great upheaval now in train for the British university system on this example, nobody could have accused it of inconsistency. In the modern state, and with the need of skilled manpower exceeding the natural supply, it would have been necessary that the numbers of intending students (of both genders) should not be determined solely by the inclinations and fortunes of parents. Curiously, the British government could have arranged for that outcome by administrative action alone, by requiring that universities should charge their students economic tuition fees (with the understanding that local authorities would, as now, meet the costs if students' families could not). The advantages, to a zealous and radical government, would have seemed to be irresistible: the need for a whole layer of bureaucracy would have disappeared, while individual universities would have been given the most powerful incentive for standing on their own feet — the knowledge that failure to attract students would rob them of income, perhaps even of continued existence.

Battle

In the middle of a bruising parliamentary battle over its education bill, the British government must be wishing that it had followed some such course. There would have been political difficulties, it is true, not least those suggested by the discovery in other recent parliamentary scrapes that the British middle classes, while relishing increased prosperity and lower taxes, are also infected by what is called the "dependency culture" of the welfare state and therefore unwilling to follow the Bologna example. But those would have been less serious than the troubles that will follow from the decision to impose what amounts to central direction of the British university system. So much is splendidly typified by the government's defeat (see page 289) in the House of Lords last Thursday on the issue of academic freedom.

The circumstances are odd, to say the least. The Education Reform Bill, first published six months ago, embodies the ambition of the past two years that contracts between universities and their academics should not in future include the promise by universities of life-long tenure for university academics. The

strength of the government's feeling on the issue seems to derive from its conviction that university teachers lead an over-protected life. The universities have reluctantly agreed to bow to the government's wishes but have asked that the bill should include provisions that academics cannot be dismissed from their jobs because their opinions are eccentric, unfashionable or simply unpopular. Last Thursday's case was launched by the Chancellor of the University of Oxford, Lord Jenkins. The government appeared to concede that there is a case to answer. Lord Mackay, the distinguished Scottish judge appointed Lord Chancellor earlier in the year, promised that the government would introduce a suitable form of words if it could think of one, but confessed that the task would not be simple. So the House of Lords voted against the government. Now there are two choices: either the government finds an appropriate form of words in the next few days, or its bill will be delayed for perhaps six months (the limit of the constitutional power of the House of Lords).

Sympathy

What the government should understand is that it would not be in this embarrassing position if it had not decided, while privatizing most state-owned industries, to take closer central control of the universities it supports. Everybody will sympathize with Lord Mackay's difficulty in drafting suitable forms of words. Under the new regime, it will be possible for academics to be dismissed for incompetence or if their services are no longer needed (as well as for the traditional and more venal reasons), but how can it be arranged in legislation that these grounds for dismissal will not cloak more sinister motives? How, for that matter, can it be arranged that the still non-existent Universities' Funding Council is not required by its powers of direction to close university departments that are collectively unpopular? And what form of words, binding on universities and those who will be running them, can ensure that British academics are never deterred from doing, saying and believing as they see fit by the fear that they may lose their jobs?

The difficulty could have been anticipated. That it was not probably derives from the general failure to understand why academic freedom matters. But there is a sense, maddening to civil servants and governments though it may be, in which it is part of the proper role of a good university to be socially subversive (in the widest sense). Where should we be if historians were not regularly undermining chauvinistic myths, literary people teaching James Joyce as well as Shakespeare, economists explaining why government economic policies will not succeed or people with technical knowledge explaining why successful industries may nevertheless be bad for people's health or why too much carbon dioxide in the atmosphere may be bad for all of us? So what should happen to, say, a political scientist who extols the effectiveness of terrorism, a geneticist who marshals evidence that intelligence is inherited or technical people opposed to central tenets of government policy (as many US academics have opposed the Strategic Defense Initiative)?

No formal definition can distinguish between eccentricity and seditious intent (let alone say what students should be allowed to hear). But universities have means of doing just this, by means



of the process of academic disputation by which university departments are made the contentious places. Ultimately, only an academic's close colleagues can accurately tell the difference between academic eccentricity and misguided incompetence. If the British government's complaint had been that academics have been less than zealous in making these judgements of their colleagues' good sense, that might have carried weight. But that it should think of creating circumstances in which the careers of apparently eccentric academics should be terminable by outsiders, with interests different from their own, is a principle than should not be conceded, whatever forms of words Lord Mackay produces. Meanwhile, the British government may note regretfully, one benefit of what may be called the Bologna solution is that it would have fallen to others than itself to solve the problem. Indeed, as the legend of Bologna shows, the secret of a university's survival is the particular blend of tough-mindedness and tolerance that marks out the better institutions. □

Too much for food

An international comparison shows that all industrial nations spend too much on subsidizing food.

AFTER years of intercontinental squabbling between those industrialized nations who also behave as if it were vital to their national interests that they should maximize their agricultural production, these partners in crime are beginning to get to grips, within the framework of the Organisation for Economic Co-operation and Development (OECD), with the enormity of their transgressions of commonsense. Their chief incentive is the certainty that within the next three to five years, as the new round of negotiations under the General Agreement on Tariffs and Trade gets under way, they will be further exposed to external pressure to behave sensibly. But there is also the continuing tripartite squabble in which the three chief trading blocks repeatedly accuse each other of over-subsidizing domestic agriculture with such ferocity that serious damage is likely to be done to their other interests.

None of this implies that cut-throat competitive farm subsidies (decreed by governments, but financed by consumers either as taxes or as higher prices) are about to end. That would be too much to ask for. What the industrial farm producers are now about is merely to investigate the enormity of their subsidies to agricultural producers. The task is not a simple one, given the multiplicity of the means by which subsidies can be paid, so that OECD's figures are best used comparatively. But there can be no question that the subsidies are enormous — a calculated average of 47 per cent of the value of agricultural production in 1986. Among the members of OECD, the percentage subsidy ranges widely, from 15 per cent in Australia to 75 per cent in Japan. The European Communities, at 50 per cent, are just above the middle of the range, with the United States way back in the field at 35 per cent.

Several questions arise, of which one of the chief is that of whether it can make sense for the industrialized farm-surplus nations vigorously to attack each other for subsidizing farm production when, evidently, they are all doing more or less the same. Second, there is the serious question whether they are now so wedded to the practice that they may not be able to give it up; the plight of farming communities caused by the ending of subsidies would not be the less harrowing (or politically influential) because the cause was just. Probably the best that can come out of GATT will be an understanding that subsidies will at least be contained for a time. The real losers in this controversy are not the farming communities but the still larger company of food-consumers throughout the world, to whom the cost of food is getting on for twice what it needs to be, not to mention the developing nations of the world, for many of which agricultural production would be a palatable substitute for the poverty in which they now live. □

Delayed START

A superpower deal on strategic missiles will not now come next week, but that does not spell tragedy.

NEXT week's meeting in Moscow between President Ronald Reagan and Mr Mikhail Gorbachev seems unlikely to see the signing of a treaty to reduce numbers of strategic nuclear missiles by a half, but that is understandable and forgivable. Indeed, the two men will be lucky if they can announce that last year's treaty on intermediate missiles (INF) has been ratified. On past weeks' form, the US Senate will still be debating the issue when Reagan leaves to catch his plane. Yet neither is blaming the other because there is no strategic treaty, even though both were saying just six months ago in Washington that they were aiming at just that. What has happened since then to make them so tolerant of the delay, and of each other?

Gorbachev has other things on his mind. The party conference arranged for Moscow at the end of June, at which the policy of *perestroika* will be affirmed, denied or possibly watered down, is in every way more important — for the United States as well as the Soviet Union. Naturally, it might even have helped to make the conference go well if there had been a strategic treaty to sign, but the seriousness of the outstanding issues and the delay in the Senate over ratification of INF may have persuaded Gorbachev that it would be preferable, instead, to deal with Reagan's successor.

What can be expected to happen after all the dust has settled? The INF treaty, while it helps to strike a safer strategic balance in Europe and elsewhere, is a huge legal agreement stuffed with carefully negotiated details, which is one reason why the US Senate ratification has been so slow. But the Senate is not on that account to blame. The successful negotiation of INF last year broke new ground in the design of treaties of this kind in the acceptance by both sides of the principle that verification should be thorough and comprehensive.

That is as it should be, but the result has been the string of questions arising in the past few weeks. How can we tell what kinds of missile components are in canisters we are not allowed to open? Can foreign nationals exercising their treaty rights to inspect parts of a factory once used to manufacture of a forbidden missile also look at other parts, supposedly concerned with a missile model not forbidden by INF? Especially because changing circumstances will complicate the interpretation of the language that now constitutes the treaty, there is something in the view that such a complicated treaty is a recipe for trouble. Yet the draft of the strategic treaty (affectionately called START) is already 300 pages long, consisting mostly of references to questions yet to be negotiated. May it already be too long to be workable, except by teams of international lawyers?

That is one of the reasons why the arms control process needs to be thought through again. The present climate of mutual regard (trust would still be too strong a word) between Washington and Moscow, much changed even in the past year, suggests that it would now be best to concentrate not on the grand goals but on the smaller issues that might be simply as well as quickly settled. Given that INF will have its chief effects on the European balance, but that a little time will have to pass before shorter-range nuclear weapons can also be regulated, should not more energy go into the regulation of conventional forces in Europe? May this not also be a time for hoping that the comprehensive test-ban treaty can be made to come to life again, perhaps in time celebrate the centenary of when it was last nearly signed? As for strategic arms themselves, a treaty would evidently be a splendid achievement, but it would be even better if these weapons systems came to seem so much an irrelevance that their formal banishment by treaty seemed unnecessary. It will be a great help if those meeting in Moscow next week remember that arms control is not an end in itself, but merely one means to another. □

Legasov's indictment of Chernobyl management

■ Senior atomic scientist commits suicide

■ "Lessons" still to be learned

London

THE Chernobyl disaster was the "apothecosis and peak of the economic mismanagement in our country over decades", dictated Academician Valerii Legasov, shortly before he took his own life last month. Legasov, chief deputy director of the Kurchatov nuclear energy institute, had been one of the first scientists to visit Chernobyl after the accident, and in September 1986 headed the Soviet team at the post-accident conference convened by the International Atomic Energy Agency (IAEA) in Vienna.

Last year, *Pravda* asked Legasov for his personal recollections of the accident, of the clean-up and for his thoughts on the future of nuclear power. His notes appeared in *Pravda* last week, filling almost a page and a half — the sort of coverage normally devoted only to major party declarations.

Legasov, whom *Pravda* describes as "simultaneously a Don Quixote and a Joan of Arc", was clearly worried about the safety of nuclear reactors long before the accident. The notes cite the practice in the United States of making "systematic assessments of the probability of every type of accident", saying that there was not a single team in the Soviet Union with "even the slightest competence" to pose and solve such problems.

Even the best of the scientists working on nuclear safety, V.A. Sidorenko, concentrated, Legasov said, on organizational methods of dealing with incidents and better documentation systems for power station constructors and operators. The top personnel of the nuclear industry became older and more set in their ways, new approaches were unwelcome and suggestions that there should be a new approach to safety met with increasing hostility. Reactor builders did indeed consider the RBMK (the reactor type installed at Chernobyl) unsatisfactory, but only because of its high cost and fuel consumption, not for reasons of safety.

Even the directors of nuclear power stations seemed oblivious of danger. "What are you so worried about?", one of them asked Legasov. "An atomic reactor is only a samovar, it is much simpler than a thermal station, we have experienced personnel and nothing is going to happen, never!"

At 1.24 a.m. on 26 April 1986, however, the impossible happened. By the afternoon of the same day, Legasov was in

Kiev, where he found the leaders of the Ukrainian SSR in a state of alarm but with no precise information other than that "things are bad". There was considerable confusion over such vital matters as respirators, and radiation meters (both of which were in short supply) and there were no leaflets available for distribution to the population setting out in simple terms the dos and don'ts of radiation hygiene. The fire-fighting operations were equally badly organized, it appears. Legasov notes that for lack of adequate ground support, helicopter pilots had themselves to load sandbags before flying them in to dump on the burning reactor.

The immediate cause of the Chernobyl accident, the government commission found, was the decision to go ahead with



Valerii Legasov as pictured in *Pravda* — another Chernobyl victim?

an experiment on the rotors, under conditions when it was no longer safe, and also the unauthorized disconnection of the safety systems to make this feasible. But, Legasov concluded, when one considered the whole chain of circumstances, the mistakes and attitudes which surrounded the accident, it was impossible to point to one particular person or decision as the guilty party or the cause of the disaster.

"The lessons of Chernobyl have still not been analysed to the end", Legasov told *Pravda* just before he took his life.

Whether this statement had any connection with his death is unclear — *Pravda* revealed the fact of his suicide not the motive. But his final article, to which *Pravda* gave the title "It is my duty to speak out", may well lead to the "lessons" being re-examined.

Vera Rich

IAEA's verdict on Chernobyl

Vienna

CUCUMBERS grown in contaminated soil from the Chernobyl area were sampled by a delegation from the International Atomic Energy Agency (IAEA) visiting the Soviet power plant last week. The visit coincided with a conference in Kiev on medical aspects of the Chernobyl accident.

IAEA director-general Hans Blix and his companions ate cucumbers and radishes that had been grown in contaminated soil in a greenhouse in Pripyat, a few kilometres from Chernobyl. Researchers there are testing a variety of plants to determine how much radioactive caesium and other elements they take up. They are also seeking useful mutations resulting from the radiation. Blix said that the cucumbers and radishes were among plants that did not absorb much caesium.

According to one IAEA official, the conference was told that, on the basis of a study of 30,000 people living in the area, no adverse health effects on the general population had been attributed to the radiation. It is known that 31 people died in the accident and more than 200 suffered radiation sickness. The recently expanded Centre for Radiation Medicine in the suburbs of Kiev plans to continue its search for long-term effects of the radiation on 600,000 people who were exposed.

The IAEA team reported several changes since its last visit in January 1987. Another reactor had been returned to service, immediately adjacent to the concrete-encased reactor, in which the temperature has dropped from 106°C to 36°C. Large-scale soil decontamination continues within a 30 km radius.

Plans to build two more reactors at Chernobyl have been shelved, said Blix. More than 1,000 elderly villagers have returned illegally to 20 villages in the 30-km zone, but the authorities will apparently allow them to stay. The Soviet government seems to consider the residual radiation less of a risk to these people's health than the continuing separation from their homes. Although there are still a few hot spots, most of the area within 10 to 30 km from the reactor has returned to normal levels of radioactivity.

A recent and highly unusual report in *Pravda*, had criticized the managers at the plant for hiring unqualified workers; Blix said that the Soviets had confirmed the report but that plant operators had not been affected. The unqualified personnel had been attracted by the high pay offered and the payrolls had become "swollen" with unskilled workers, said Blix.

Steven Dickman

The developing countries are seeking a bigger say in IAEA affairs. See page 288.

Met. Office and defence research lose out in spending proposals

London

THE British government is to spend £2,272 on military research and development (R&D) in the current financial year, of which £403 million will go on research. The figures for last year were £2,346 million and £401 million respectively. The decrease in spending on total R&D and the decline in real terms in research expenditure is in keeping with the government's recently declared policy of lessening the share of the nation's research resources devoted to defence. Even so, the government is shying from instituting substantial reductions and shifting the balance to the civil sector. The statement of the government's defence spending estimates, released last week, says "we continue to pay very close attention to the real level of defence research and development expenditure, the aim being to achieve a gradual reduction over the next decade . . . there will, however, be no abrupt change; the Ministry of Defence will continue to fund R&D on a substantial scale".

The statement will provide little comfort for the already squeezed defence research establishments, dogged by shortages of manpower and resources (see *Nature* 332, 479; 1988). The six non-nuclear research establishments were earlier this year given clearance to sell their services in the market-place and are likely to be early candidates for agency status under recent government proposals to give greater autonomy to civil service establishments.

For the Meteorological Office, part of the Air Force Department, the statement's three short paragraphs give little inkling of the serious financial problems the office is experiencing. But for an independent inquiry into the Met. Office's failure to give adequate warnings about the 'great storm' of 15-16 October last year, many of the difficulties would remain unrecognized.

Accepting the recommendations, the government made clear at the outset that no more money would be available to meet them. Although this year's defence estimate does not give specific figures for the Met. Office, its budget is likely to be pegged at around last year's figure of £81 million. Of this, £23.6 million was obtained through the office's commercial activities. The Met. Office is under instruction to identify economies amounting to some 7 per cent over the next three years. Serious doubts are being raised about the office's ability to provide an effective service faced with such pressures of economy. Over the past 12 years, total staff numbers have fallen by nearly 40 per cent to the

present 2,500. Plans to give the Met. Office executive agency status by 1990 are well under way and have in general been welcomed by staff, who accept that the key to the office's future lies in the expansion of its commercial activities.

On participation in the US Strategic Defense Initiative (SDI) research programme, the statement remains subdued. By February this year, some 65 contracts had been placed in the United Kingdom, to a value of \$60 million. The ministry had earlier predicted that by the end of 1987 contracts worth \$100 million would have been placed in the United Kingdom. Contracts worth \$10 million have been awarded to 17 higher academic research bodies. The ministry concedes that the SDI work "has been gained in a climate of competition . . . that has turned out to be more severe than predicted".

Simon Hadlington

North Sea project lacks the cash

As final preparations are made for the experimental stage of Britain's five-year study of the chemical, biological and sediment processes in the North Sea, the project's administrators are still trying to make up a budget shortfall of more than £2 million. The North Sea Project is being carried out principally by the Natural Environment Research Council (NERC), which is providing £7 million of the £10.6 million needed. Separate but related studies by the Ministry of Agriculture, Fisheries and Food will contribute a further £1.3 million. The Department of the Environment (DoE) has declined to contribute substantial resources to the project as a whole, but has been persuaded to put relatively small amounts into specific parts of the programme.

The aim of the project is to formulate three-dimensional hydrodynamical and transport models for predicting water quality. From early August the research vessel *Challenger* will spend 15 months plying the sea along a set track gathering data, which will be processed on shore.

Project organizers say that the shortfall of funds will not severely compromise the experiments, but will leave little room for flexibility — for example there are no contingency funds available to replace loss of equipment, an almost inevitable hazard in shipboard studies. The hope is that the Department of Energy may be persuaded to take a financial stake, and that private companies with an interest in the North Sea may supply funds.

Simon Hadlington

SDI simulations under fire

Washington

A GROUP of computer scientists has fired a broadside at the Strategic Defense Initiative's (SDI) National Test Bed being built at Colorado Springs, Colorado. A report by Computer Professionals for Social Responsibility (CPSR) last week says that the idea that the test bed can determine whether SDI will function is "based on a fundamental misunderstanding of the value of computer simulation."

Undeterred by this criticism, the SDI Organization claims that the test bed can be used to evaluate the various hardware and software components that will make up an effective ballistic missile defence.

Martin Marietta last January signed a contract worth \$508 million over 5 years to design, install and operate the National Test Bed (NTB), which consists of a network of test sites scattered around the United States, coordinated through the National Test Facility in Colorado. NTB will use supercomputers and electronic test equipment to provide a simulation environment for SDI's component systems, as well as to verify strategies for overall battle management capabilities.

But CPSR says it is impossible to develop meaningful simulations of a space-based missile defence, because performance specifications cannot be validated against real data derived from experience. Even assuming all the parameters of an intended defence system can be accurately represented in models, it is not possible to include all the ways in which the system might be subverted.

The report argues that exercises using hypothetical constructs of potentially real situations are not simulations but "hypothetical presentations that may resemble some chain of events in the real world", only providing an accurate picture of the world dreamed up by their creators.

David Parnas, a former SDI adviser from Queen's University in Ontario, Canada, says NTB will be nothing more than "the world's largest arcade game parlour", and will offer no true insight into the reliability of an SDI system.

But Danny Cohen, an adviser to the SDI Organization from the University of Southern California, says CPSR's criticisms are ideologically motivated. Cohen maintains that any defensive system must make assumptions about the attack it is going to face. Although it is true that US planners will have doubts about Soviet offensive capabilities, he points out that Soviet planners will also have doubts about SDI's effectiveness. This, Cohen maintains, will add to deterrence, and limit the likelihood that a war will start.

Joseph Palca

New Hungarian science union to face post-Kadar era

London

HUNGARY'S new Democratic Union of Scientific Workers was formally established on 14 May. Following a postal survey of Hungary's 2,000 research establishments (see *Nature* 352, 385; 1988) it was decided to set up a single union for all scientists and academic researchers. This cuts across Hungary's existing union structure, which makes union membership dependent on workplace, and so spreads the country's 77,000 scientists and academic researchers between 16 unions. The relationship of the new body with the existing National Trade Union Council (SZOT) is at present undefined.

During the past year, hundreds of thousands of Hungarians have left the existing union movement. The scientists' discontent, however, goes back a number of years, as, according to Csabo Oeri, one of the spokespersons for the new union, the scientists have felt their personal and professional situation to have been consistently "devalued". Unlike the SZOT unions, the new organization, being independent of the workplace, will be able to take care of scientists who lose their jobs. More than 1,000 scientists and academic researchers have already joined.

The scientists' initiative is likely to inspire other professions to set up independent unions. Gyoergy Kerekcs,

another spokesman of the new union and an active party member, predicts that in two years there will be several dozen profession-based unions. Not surprisingly, there is some reticence, even in the academic establishment, towards the new body. (The founding meeting could not be held on university or academy premises, but was held in a metal-workers club.)

The founders of the Democratic Union of Scientific Workers stress, however, that they do not want to challenge the state nor to become a political force in the style of Poland's 'Solidarity'. What they want, they say, is to stand up for the rights of the country's scientists and scholars, "with a minimum of bureaucracy and a maximum of democracy".

Vera Rich

■ Although Hungary's new party leader, Karoly Grosz, has made clear his commitment to the "democratization" of Hungarian life, his attitude to the concept of "alternative" unions is somewhat negative — to judge from an interview he gave the government newspaper *Magyar Hirlap* in April. During his visit to London, three weeks ago, Mr Grosz was asked to enlarge on this apparent dichotomy, he said that he respected other people's opinions on multi-party systems, and independent trade unions and youth organizations — and that he had the right to expect other people to respect his views too. □

Labour outlines policy for more industry-led research

London

A FUTURE Labour government in Britain would base its industrial strategy on greater public and private investment in science, research and development, the party's trade and industry spokesman said last week. Delivering the last in a series of speeches on the party's recipe for industrial and economic growth, Bryan Gould insisted that Britain needs an "industrial and technological revolution" to remedy "our dismal record on science".

In the first instance, investment in basic science must increase, and greater account must be taken of the rapidly rising costs of vital equipment. The demand by the pressure group Save British Science for an immediate injection of £100 million followed by a 10 per cent increase in civil research spending over the next five years "is a modest starting point" Gould says.

As part of its "medium-term industrial strategy" a Labour government would launch several initiatives to encourage greater investment by industry in research and development. Companies would be

obliged to disclose their research and development expenditure. "We should seek to change the accounting practice which regards R&D spending as a drain on liquidity rather than an investment in an asset", Gould says.

Such moves would not, however, make many inroads into the £3,000 million increase in industry-backed civil research and development deemed necessary by Gould, who points out that even the present investment is dominated by just three industries — electronics, aerospace and chemicals — which spend about ten times as much on research and development per employee as the rest of manufacturing.

A Labour government would therefore want to assist specific companies to increase their research spending, for example by a grant related to the increase in research and development undertaken, "although to stop all funds going to the existing high-tech firms there might need to be a ceiling put on the overall sum or the amount per employee", Gould says.

Simon Hadlington

Arizona backs insects

A \$3.2-million Center for Insect Science at the University of Arizona is one of three new projects that will receive funds from the National Science Foundation (NSF) under its biological research centres programme.

The University of Arizona centre embodies the traditional goals of entomology departments and agricultural experimental stations as well as more speculative efforts to discover new uses for insects and to use them as models for understanding mammalian systems.

"We think it's a vastly neglected group of animals", says John H. Law, the centre's acting director. Nicholas J. Strausfeld, a professor of neurobiology and a member of the centre, says that insects are potentially a guide to the development and functional organization of complex nervous systems. The NSF funds will be used over the next four years to train 63 undergraduate and graduate students and postdoctoral researchers in insect biological sciences and to support short-term trainees.

The new centre expects to get \$1.7 million from the \$6 million-dollar NSF programme. The University of California at Berkeley and Johns Hopkins University are the two others, out of the sixteen finalists, to get grants from the research centre's programme.

E.P.

Nicotine addiction

NICOTINE is the drug in tobacco that causes addiction. That is the conclusion of the twentieth US Public Health Service report on the health consequences of smoking released last week. The 618-page report chronicles the pharmacological and behavioural basis for nicotine addiction, showing that it acts in a way similar to that of cocaine and heroin.

Surgeon General C. Everett Koop argues that prevention of tobacco use should be included along with prevention of illicit drugs use in school health education programmes. He further suggests that policy makers should question whether tobacco sales outlets should be licensed just as are alcohol sales outlets.

J.P.

Demolition halted

BULLDOZER operations on the site of a Byzantine cemetery in the southern Crimea (see *Nature* 333, 198; 1988) have now been halted, according to a message just received from the Leningrad Institute of Archaeology of the Soviet Academy of Sciences. The site was being cleared — in defiance of Soviet regulations on archaeological finds — so that a rest home could be built for the Soviet Academy of Sciences. This paradoxical — and illegal — situation seems to have resulted from the attempts of a local functionary of the academy to build himself a personal power base. He has now been formally reprimanded.

V.R.

Developing countries plan to displace Blix at IAEA

Vienna

STORM clouds have appeared on the horizon at the International Atomic Energy Agency (IAEA). The agency had emerged as a poised leader in the international response to the 1986 Chernobyl nuclear accident and had pleased its sponsors in the developed world. But the time of tranquillity may soon come to an end.

The Group of 77 (G77), which represents developing countries in international agencies such as the United Nations (UN) and the IAEA, announced its dissatisfaction earlier this spring with the proposed agency budget for 1989 and with Hans Blix, the agency's director-general. G77 has hinted that it will announce its own candidate for director-general before Blix's second term ends in 1989.

The Vienna head of G77, Mexican Ambassador Francisco Cuevas Cancino, denounced a proposed increase in the IAEA's \$50 million annual budget for nuclear safeguards. The safeguards increase would require cuts in IAEA programmes in developing countries.

According to the news agency IPS, Cuevas Cancino declared that the agency budget lacks "audacity" and "multilaterality." The criticism reflects a continuing struggle within the IAEA between states that have nuclear technology and those that do not. A few developed countries contribute the lion's share to the agency

A proposed increase of \$2 million in the safeguards budget for 1989 brought on the current disagreement. Some of the money derived from efficiency gains in non-safeguards areas. The increase pushed the safeguards budget to slightly more than one-third of the IAEA's total budget.

Blix last week compared his budget task to "squaring the circle". The work load of the agency has been increasing, without a corresponding increase in the safeguards budget, which has stayed the same for the past four years. At the same time, developing countries demand that the balance between safeguards and other programmes should remain the same.

Blix tried to accomplish the impossible by asking developed countries to provide

more money in the form of voluntary contributions. His request was rejected.

The latest form of Blix's budget, in which a few posts have been shifted out of the safeguards area, will be voted on at the June board of governors meeting in Vienna. Blix predicts that his compromise proposal will be accepted by the board.

Challenging Blix for the post of director general will not be as easy for G77 as challenging the budget. Despite a 1985 understanding that Blix would be replaced by a candidate from G77 in 1989, IAEA sources consider it unlikely that the G77 can unite behind a single person. There are some nuclear experts from developing countries who would be available and eager to take the job.

Blix will announce his decision to run before the IAEA general conference in September 1988. The election will take place at the general conference in September 1989.

Steven Dickman

Where does India find heavy water for Candu reactors?

New Delhi

A REPORT giving weight to international press reports alleging that India might have acquired heavy water illegally — outside the provisions of the Nuclear Non-proliferation Treaty — has just been released by the comptroller and auditor general (CAG) of India. The report has shattered official claims about self-sufficiency in heavy water.

The audit report infers that India produced less than 200 tonnes of heavy water in the period 1978–86, at a time when three Candu reactors were brought into service, requiring a total of 578 tonnes of heavy water. India claimed that the heavy water used in these reactors was its own.

The report has received wide publicity in the wake of a new allegation that 15 tonnes of heavy water found missing in a Norwegian plant in 1983 might have ended up in India. The government's denial has been weakened by the audit report, which suggests that the stock of indigenous heavy water fell far short of the quantity needed to charge the two reactors of the Madras Atomic Power Project (MAPP), and the Dhruva research reactor in Bombay. MAPP-I, commissioned in July 1983, and MAPP-II, commissioned in 1985, each required 250 tonnes; Dhruva needed 78 tonnes.

The report says that the Tuticorin heavy water plant, commissioned in July 1978, was the best among all operational plants, and yet its average annual production for eight years was only 20 per cent of its rated capacity of 71 tonnes. Between 1978 and 1986, this plant produced a total of 114 tonnes — "60 per cent of the indigenous heavy water".

An official of the CAG office says that it took a "circuitous" approach in stating the cumulative production in keeping with the policy of the Department of Atomic Energy (DAE) of secrecy over exact production figures of individual plants. The DAE declared such information as classified soon after commissioning MAPP-I, apparently to ward off speculation about the source of heavy water.

The DAE has admitted that in July 1982, its stock consisted of 45 tonnes of fresh heavy water and 105 tonnes of used heavy water that required upgrading. But it has not explained how it managed to make up the shortfall of 428 tonnes between 1983 and 1985. Only three heavy water plants — at Nangal, Baroda and Tuticorin — were then operating and, if the CAG report is correct, they would have produced at best 80 tonnes.

US reports two years ago that India had made up its shortage of heavy water by importing it from China and by siphoning it off from the disused Rajasthan atomic station were denied by the Indian government. The Soviet Union has officially stated that it has so far supplied 450 tonnes of heavy water to India under the safeguards of the anti-proliferation treaty. This, according to DAE, has been used in the two reactors of the Rajasthan heavy atomic power station.

India has six operational heavy water plants, with a total capacity of 350 tonnes, and the seventh is under construction. The present stock is not known but DAE would need 250 tonnes for the first unit of the Narora atomic power project scheduled to become critical before the end of the year.

K.S. Jayaraman



Hans Blix — director-general of the IAEA.

budget — G77 countries pay only 2 per cent of the safeguards budget, which is used to administer the 1971 non-proliferation treaty to prevent the diversion of nuclear material for use in weapons.

But many developing nations benefit from IAEA programmes in nuclear medicine and agriculture. Thus, in addition to serving as a nuclear inspection squad for the developed world, the IAEA is also a conduit for the transfer of technology to less developed countries.

Lords defeat for government on academic freedom clause

London

BRITISH academics are savouring a second significant victory in defence of university autonomy. Last week in the House of Lords, the government failed to defeat an opposition amendment to the education reform bill giving statutory protection to academic freedom. In its original form, the bill, which is likely to become law in the summer, proposes that government-appointed parliamentary commissioners will rewrite university charters and statutes to abolish tenure and to enable staff to be dismissed "by reason of redundancy" or "for good cause". The implications of such a provision were seen as particularly insidious when considered in relation to other proposals (since modified) giving the government powers to direct the funding of individual departments within institutions. Despite sustained protest, much of it from the government's own supporters, the government has argued that it is impossible to draft a safeguard of academic freedom that would not be open to abuse.

Last week's successful amendment, put forward by Lord Jenkins of Hillhead, chancellor of the University of Oxford, would ensure that "academic staff have freedom within the law to question and test received wisdom, and to put forward new ideas and controversial and unpopular opinions, without placing themselves in jeopardy of losing their jobs or privileges that they may have at their institutions". The peers voted by 152 to 126 in favour of the amendment. Although the government is able technically to throw out the amendment when the bill returns to the House of Commons, the majority is seen as sufficiently convincing to make such a rejection unlikely. The wording of the amendment was originally put forward by the Committee of Vice-Chancellors and Principals several months ago, and was in essence accepted by the Education Secretary Kenneth Baker. Ironically, Baker is now said to be unhappy with the wording of the amendment.

The government has clearly failed to gauge the strength of feeling on its attempts to interfere with the running of the universities. It had ample opportunity to avoid the embarrassment of a defeat in the House of Lords. For more than six months the vice-chancellors have been meeting with government representatives to try to resolve the question of academic freedom, but with no sign of progress. The government's view has been that a comprehensive grievance procedure for dismissed academics would be sufficient to safeguard an individual's intellectual freedom.

The concessions the government will now be forced to make in response to last week's vote in the Lords will be the second major climbdown on the higher education aspects of the bill. Three months ago, when the bill was being considered by the Commons standing committee on education, Baker bowed to pressure and agreed to a number of important concessions on the government's powers to intercede in the financing and management of higher education (see *Nature* 331, 645; 1988).

Simon Hadlington

More ECUs to spend

THE Council of the European Communities last month approved two projects — SCIENCE and DELTA — that aim to promote exchange and training facilities for European researchers. Carrying on from the previous Stimulation Programme, SCIENCE aims to create a network of exchange and co-operation between European laboratories and to facilitate visits by young researchers to other European laboratories. With a budget of 167 million ECU (1 ECU = £0.69) for the period 1988–92, the programme will help over 8,000 researchers working in the exact and natural sciences.

DELTA, which will be complementary to the existing COMETT educational programme, is to develop and exploit new educational technologies, based around information technology, telecommunications and broadcasting. With a preliminary budget of 20 million ECU for an 18-month trial period, DELTA aims to encourage the development of norms in both technology and curriculum for Europe-wide educational programmes.

P.C.

Quick is beautiful

ALTHOUGH giant budgets and years of planning have recently been the norm for most on-orbit projects of the US National Aeronautics and Space Administration (NASA), there is a new theme being heard around the agency: 'Quick is beautiful'. 'QIB', as the philosophy is known at NASA, would be especially appropriate for the space station, according to a study group assigned to look into the topic. Payloads that require minimal amounts of power, communications capacity and/or human attention would be extremely useful for graduate students in space science or small companies anxious for 'proof of concept' data. The study group recommends that NASA create a management structure for QIB projects, and that five per cent of total space station resources be devoted to QIB research.

J.P.

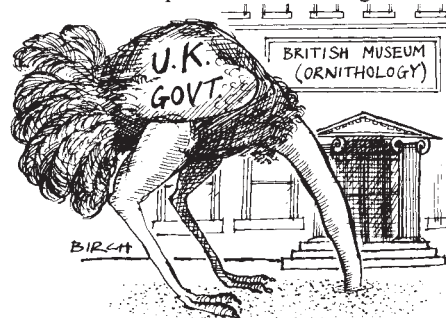
Natural History Museum in a decline?

London

CONCERN about the future of research in Britain's principal museum for the natural sciences has been rekindled with the announcement that active research is to cease in three areas. Declining government support in recent years has resulted in financial stringency and low morale among researchers at the British Museum (Natural History). Last month, following an external peer review, the museum authorities announced that research at the sub-department of ornithology, at Tring in Hertfordshire, was to finish, with the two individuals affected being relocated elsewhere in the museum. Research into arachnids and coelenterates is also to be discontinued. Again, individuals affected will be relocated.

The British Ornithologists' Union has deplored the decision to pull out of bird research. It points out that since 1980 the number of staff at the ornithology sub-department has been reduced from eleven to six, and that once the new proposals are implemented, together with an impending retirement, all that will remain will be three junior staff "lacking any qualifications to carry on the other functions of the sub-department."

The decision to cut back on some areas of research comes at a time when ministers and museum officials are considering a confidential report from the Programme



of Policy Research in Science and Technology (PREST), which draws attention to the level of despondency among the scientific staff and concludes that major changes in the museum's policy will be necessary to avoid long-term decline.

Museum officials are keen to point out that no direct cost savings will result from the latest moves, and that no redundancies have been made. The decision was taken essentially on scientific grounds, for more efficient deployment of expertise within limited resources.

Simon Hadlington

■ The debate over museum cutbacks continues — see Correspondence, page 292.

Graduate university aimed at research flexibility

Tokyo

A UNIQUE new graduate university is to be formed under the terms of a bill passed last week by the lower house of the Japanese Diet. The university, which will offer doctoral courses to scientists with masters' degrees, will introduce some flexibility to Japan's rigid university research system, increase the competition for graduate students and perhaps help to open the country to foreign scientists.

The idea of creating the "General Research Graduate University" was first officially put forward in 1982 by the directors-general of the national research institutes, mostly established in the 1970s as central facilities for use by the universities.

Although among the best-equipped research institutes in Japan, the institutes have not been free to enrol their own graduate students or to award degrees. Graduate students wishing to use an institute's facilities have instead been sent by outside university professors.

The new graduate university, to be jointly administered by the research institutes, will provide a supply of graduate students and award doctoral degrees. It will open in October, initially on a temporary site on the campus of the Tokyo Institute of Technology in Yokohama. The first 50 students will be accepted next April and dispersed to the participant research institutes — those already cited plus the institutes of genetics (Shizuoka) and statistical mathematics (Tokyo). There will be three departments to begin with (in physical, life and cultural science), but the organizers also hope to add a joint research department for interdisciplinary research.

Students with masters' degrees will be accepted from universities throughout Japan, but Professor Saburo Nagakura, president of the Okazaki National Research Institutes, who is in charge of setting up the university, is particularly keen to accept foreign graduate students; he has already discussed the possibility of student exchange with a British university.

Nagakura hopes that the new graduate university will also introduce some flexibility into a system in which lateral movements of students, teachers and researchers is notoriously small.

But some of the national laboratories have decided not to join the new network, continuing instead to recruit graduate students from their mother universities (in these cases, Tokyo University).

Indeed, Tokyo University has plans to strengthen its graduate research programme by fusing the undergraduate and graduate schools, offering a six-year

master's course. But this proposal is opposed by some of the university's research institutes, which regard it as an attempt to 'steal' their graduate students (see *Nature* 330, 597; 1987).

Behind all these moves lies a battle among Japan's professors to gain access to outstanding graduate students. But

Nagakura says that the new graduate university is not intended to develop into a 'tug of war' over graduate students, a view echoed by Professor Akito Arima, vice-president of Tokyo University, who is promoting the university's graduate school reform.

With an intake of only 50 students a year, the new graduate university will have a student population of only 150–200, so that there will be plenty of students for everyone, Nagakura says.

David Swinbanks

US plan for the protection of African elephants

Washington

To reduce the slaughter of African elephants for their ivory, the US-based African Wildlife Foundation (AWF) last week called for people to stop buying worked ivory jewellery and other products. Short of a complete ban on ivory imports, AWF considers a US consumer-embargo the most effective means to cut down on the illegal poaching of African elephants.

AWF associate Cynthia Moss says poaching has cut the African elephant population by half since 1979, and adds that each year 80,000 elephants die to support the demand for worked ivory — 70,000 directly for their tusks and another 10,000 young elephants after their mothers or herd leaders have been killed.

The United States supports one-third of the world's demand for worked ivory, and an estimated 80 per cent of the ivory sold here is obtained illegally. The Convention on the International Trade in Endangered Species (CITES) lists the African elephant as a threatened species. But no effort is made to control the trade of worked ivory because its origins are so difficult to trace.

Some non-CITES African countries without elephant populations, such as Burundi, are said to buy poached ivory for export. Other African governments do nothing to control the illegal ivory trade

going on within their borders.

Moves are afoot in the United States to crack down on the illegal ivory trade. Last month the US Fish and Wildlife Service banned the import of all ivory from Burundi. Legislation has also been introduced to ban the import of all elephant products to the United States.

But most wildlife conservation organizations and ivory trade experts support a combination of reducing the demand for ivory products and tighter control of ivory imports to slow down the poaching of elephants. Many fear that a strict ban on imports would simply force the illegal ivory trade further underground, and prevent African countries from tapping one of their most important national resources once the elephant populations recover.

The US House of Representatives Committee on Merchant Marine and Fisheries is putting together legislation that may offer an alternative. The draft proposes a five-year moratorium on imports from only those African countries unable to control their ivory trade. There would also be a US tax on sales of ivory products to be channelled into an African Wildlife Conservation Fund. The draft would further ban imports of worked ivory from countries or other intermediaries purchasing ivory from countries that encourage poaching.

Carol Ezzell

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

African elephants — a species in danger.

New tests for spotting AIDS

Berkeley

A FAR more sensitive test for the human immunodeficiency virus (HIV), the virus that causes AIDS, than any now available has been developed by Cetus Corporation, of Emeryville, California. The test should be available within a month through licensed reference laboratories in California.

The principle of the test is based on DNA hybridization to detect proviral DNA. The polymerase chain reaction (PCR), developed at Cetus, amplifies the hybridization signal and allows detection of as little as one molecule of DNA. The test should receive Food and Drug Administration approval and be available in kit form in about two years.

PCR-based diagnostic tests promise to improve the safety of the blood supply, by enabling blood banks to detect HIV in blood during the weeks immediately following infection, before antibodies are present. The tests will also be useful as clinical tools, for studying the progress of viral infection and monitoring the effectiveness of experimental drugs. The sensitivity of the test, combined with internal controls, give it a very low rate of either false positives or negatives.

Cetus was awarded a broad patent on PCR in July 1987, and has filed some 20 further patent applications on refinements and automation of the process. John Sninsky, of the company's department of diagnostics research, says the first clinical kits will cost about \$300, making them cost-competitive with the current viral-culture means of detecting the virus. Further simplifications will be necessary before the test can be made simple and inexpensive enough for use by blood banks, he said.

Researchers at the City of Hope National Medical Center in Duarte, California, recently threw their hat into the HIV diagnostics ring, but will be hard pressed to catch up with Cetus. City of Hope's PCR test, developed by John Rossi, differs from the Cetus test in that it detects RNA rather than DNA. But the RNA test has so far been tried only on the blood of AIDS and AIDS-related complex (ARC) patients, and it is a matter of debate whether it will be more sensitive during the early, latent stages of the disease, when the DNA provirus, incorporated into a cell's genome, is not highly transcribed.

Rossi has been looking for pharmaceutical companies interested in developing and marketing his test, but says the response so far has been lukewarm, because it would require licensing by Cetus.

Marcia Barinaga

Attendance low puts future of ANZAAS in grave danger

Sydney

THE much-vaunted centenary of the Australian and New Zealand Association for the Advancement of Science (ANZAAS), marked by its 58th congress here last week, managed to attract the smallest number of participants (1,000 in round numbers) since figures were first kept in 1932.

Since attendance fees are a major source of funds for the association's continuing activities, its future must now be seriously in doubt after failing to attract a large paying audience in Australia's largest city.

The programme followed the standard mix at recent meetings — symposia on matters of current science and social science, plenty of media coverage and a sprinkling of overseas speakers. What has gone wrong?

There are many explanations but no solution. But the sad decline over the past century was wryly illustrated by the proud history of the association told in Professor Roy Macleod's new book *The Commonwealth of Science* released during this year's congress.

Part of the trouble of the congress must be the high cost, a registration fee of A\$60 a day, plus interstate travel where appropriate. To further complicate matters, the organizers were able to produce a detailed programme of events only the day before the congress opened.

The congress failed again to be a forum for a national debate on science policy, but there was demonstrable enthusiasm for new projects announced at ANZAAS for promoting the public understanding of science. This was the message delivered by this year's president of the corresponding British association, Sir Walter Bodmer. Dr Robert Crompton of the Australian Academy of Science announced a plan for a science and technology information service to assist the media.

One of the few exceptions, this year, to the rule that almost no original scientific work is announced at ANZAAS was the announcement by Professor Graham Johnston, the University of Sydney pharmacologist, with his colleagues David Kerr and Jennifer Ong, that steroids such as cortisol are able to modify the function of the most important inhibitory neurotransmitter γ -aminobutyric acid. It seems that adrenal steroids in very small doses (10^{-12} g) can have significant effects, pointing to the design of drugs for dealing with the problems of anxiety and stress.

Peter Pockley

Peter Pockley has covered ANZAAS congresses for Nature since 1965.

Kangaroo aid

ORTHOPAEDIC surgeons in Australia are looking to the kangaroo for help. Dr Klaus Schindhelm, assistant director of the Centre for Biomedical Engineering at the University of New South Wales in Sydney, told participants at ANZAAS that kangaroo-tail collagen is a promising new material for tendon and ligament prostheses.

Kangaroo-tail collagen has advantages over synthetic polymers used in xenografts. It is resorbed at almost the same rate as the endogenous tissue regenerates. It is also pure and of the appropriate molecular structure.

The kangaroo tail is a rich source of the collagen. A single tail from a fully grown red kangaroo may provide material for 100 tendon or 15 ligament reconstructions. The material is at present undergoing animal trials.

Rejection is not the chief problem, for collagen is highly conserved across animal species. The problem is promoting the attachment of the prosthesis to the living tissue. Schindhelm is exploring chemical modifications to do this.

Tania Ewing

Dengue fever threat

AN Australian federal government decision to cut the National Diseases Control Programme next month could result in an outbreak of a mosquito-borne disease, dengue fever, in North Queensland.

That possibility was raised by Dr Brian Kay, an entomologist from the Queensland Institute of Medical Research, speaking at ANZAAS.

The mosquito, *Aedes aegypti*, carrying the dengue fever virus, is a common inhabitant of Australian backyards. Dengue fever last broke out 26 years ago but Kay says that preventive measures have been declining since the last outbreak.

Dengue is also known as the "three-day hangover", and the typical symptoms are similar to those of a severe influenza attack. But a fatal strain of the disease, haemorrhagic dengue fever, could enter Australia from Asia. The current theory regarding virility of the disease is that antibodies raised to the classical form of dengue fever enhance the second infection by the haemorrhagic form of the virus.

"*Aedes aegypti* is easy to control and has cost the government only A\$5.5 million in the past five years", says Kay. "The rise in mosquito numbers coupled with the risk of introduction of the disease from Asia means that North Queensland is sitting on a powder keg."

Tania Ewing

Cuts at British Museum (NH)

SIR—In 1939, a proposal by the trustees of the British Museum (Natural History) to move the best ornithological collection in the world from the famous Bird Room at South Kensington to an outlying branch at Tring caused a major outcry, for reasons summarized by one of the former 'super-numerary staff' who had worked there for 30 years, the late Dr D.A. Bannerman, who wrote¹ "The study collections, which are visited and consulted by ornithologists of every nation, would automatically lose half their value through their inaccessibility . . . the contemplated move of the Bird Room from London will be greeted with dismay by 90 per cent of those with whom my work has brought me into contact".

The project was suspended for 30 years, until most of those who had objected were dead, but was then reintroduced so unobtrusively that local and international protests^{2,3} came too late to stop it. There was excessive delay over the move^{4,5}, which helped to discourage the remaining outside workers, by now treated by the growing professional staff as rather a nuisance, if not rivals, from undertaking the difficult journey to Tring, where there is now a shortage of local accommodation which has become very expensive. A large part of the egg collection was soon stolen over a period unnoticed, and the rest shuffled around⁶, since when increasing restraints have been placed upon visitors, who are no longer encouraged to assist with work on the collections.

The move to Tring might have been tolerable if what, in defiance of the informal tradition of the Bird Room, was renamed the 'Subdepartment of Ornithology' had continued to flourish, but it was one of the unspoken objections to the move that if the collections were exiled from the main museum, they might suffer disproportionately compared with other departments in hard times. Since 1980, four more or less distinguished and influential senior staff who retired have not been replaced, another is due to leave shortly, and it is said that the last two may soon be declared redundant as part of a general decision to abandon research on (of all things) cetaceans, birds, arachnids and coelenterates, leaving only three junior curators in the subdepartment to deal with the endless stream of enquiries from the public and to welcome distinguished foreign visitors.

This means the virtual end of organized research on bird systematics in Britain at a time of its explosive development as the result of the introduction of a variety of new ideas and techniques throughout the world, in which the staff of the subdepartment were beginning to play a more active role⁷. In consequence of the default

of the past management, the authors of the current vast authoritative handbook on the birds of the western Palearctic, edited in Britain⁸, have already had to go abroad for their systematics, which were provided for their predecessors through the private enterprise of a former Lord Rothschild from the same museum at Tring. It seems time that the current performance behind the turnstiles and show-cases of this national institution turned scientific Disneyland in relation to its past traditions and promises received more public scrutiny.

W. R. P. BOURNE

University of Aberdeen,
Department of Zoology,
Tillydrone Avenue,
Aberdeen AB9 2TN, UK

1. *Ibis* **82**, 183–185, 382–385, 575–577 (1940).
2. Allison, J. *The Times* 30 January 1969.
3. *Nature* **221**, 1094, 1197; **222**, 605–606; **223**, 430 (1969).
4. *Nature* **239**, 118 (1972).
5. *Br. Birds* **65**, 492 (1972).
6. *The Times* 1 May 1980.
7. Knox, A. *Br. Birds* **81**, 206–211 (1988).
8. Cramp, S. et al. (eds) *Handbook of the Birds of Europe, the Middle East and North Africa* (Oxford University Press, 1977 on, 5 vols, continuing).

SIR—The curatorial staff of the American Museum of Natural History view with alarm the effects of recent cutbacks of programme and staff at the British Museum (Natural History). The latest decisions, relegating the programmes in coelenterates, arachnids and birds to a "care and maintenance mode", has eliminated research personnel in these areas, leaving these collections of inestimable scientific worth in the care of technicians who may not, in all cases, be professionally trained biologists.

What is at stake here is nothing short of mankind's understanding of the current diversity of life. In recent years, budgeting and funding priorities, in conjunction with basic trends in biological research generally, have seen a steady concentration of high-calibre research in systematic biology increasingly in large private or government-sponsored natural history museums. As erosion of collection support and collection-based research has continued at universities, museums have struggled to take up the slack. And while systematics is vigorously pursued at relatively fewer kinds of institutions, it is also true that the intensity and calibre of such research has never been higher since the days of Linnaeus. Prominent among those institutions with the very highest quality of systematics research has been the BM(NH).

Paradoxically, while support of even our finer research institutions in systematics continues to be threatened, the public at large has seldom if ever been more aware of the need for a deep under-

standing of the diversity and connectedness of the living biota. The collections at the BM(NH), the fruits of worldwide scientific collecting for well over a hundred years, constitute one of a handful of records of the state of the living world as it is — and was — just at the onset of the current wave of ecosystem and species loss that is now taking place. Simply put, such collections are irreplaceable.

Biological collections such as those of the BM(NH) are not sentimental memorabilia of a bygone era of empire and exploration. They are our only concrete source of information about a living world that is fast disappearing. It is no extravagant luxury to maintain them — and to do so properly, under the aegis of a highly skilled and thoroughly dedicated research staff. It is, instead, a vital necessity. We urge the administrative powers that be to reconsider their decisions, and other similar plans that may be in the offing. And we urge the British public to consider whether the relatively modest sums to be saved are worth the sacrifice in commitment to preserving one of the finest sources of information about the living world.

GARETH J. NELSON

NILES ELDRIDGE

GEORGE E. HARLOW

*American Museum of Natural History,
Central Park West at 79th Street,
New York, NY 10024-5192, USA*

India and China

SIR—Contrary to the comments you published earlier (*Nature* **331**, 384; 1988), specific comparisons would reveal that India has indeed surpassed China in most fields.

The Chinese revolution stopped some three decades back, mostly through the redistribution of wealth, to provide a bowl of food, shelter and basic education, to its population at large. That too was accompanied by 50 million or more dead and mass persecutions by Chinese authorities, in contrast with the stone-age liberty that has hampered the development of what is needed most in India — a sense of national identity.

In fact, economic development was just not possible under Mao, who frowned on the notion of profit and distrusted the intellectual community to the extent that professional training all but stopped after the 1960s cultural revolution. The existence of revolutionary cadres placed by Mao at all levels of party hierarchy is one of the most difficult political obstacles for the current leadership in changing Chinese society.

The history of development is actually less than a decade old since China adopted the so-called open-door policy. Chinese in Beijing are fond of boasting about something as simple as the first luxury hotel being constructed by purely indigenous

expertise, since the international chains came up only during the past five years by massive transfers of foreign technology. Ancient equipment now lies side-by-side with the most recent acquisitions in laboratories, thanks to foreign generosity. Given the lack of adequate infrastructure for public utilities, ordnance factories had to manufacture simple appliances such as washing machines and fans to give at least a taste of comfort to less than 6 per cent of the population.

The manufacture of high-technology items such as computers and nuclear power plants still awaits adequate agreements with foreign collaborators, whereas they are taken for granted in India.

In order to replace the Soviet Union as the bastion of world communism, Mao spent enough of the gross national product to create one of the largest but most backward armies in the world. Whereas the Chinese military is struggling to modernize 30-year-old Soviet models, some of the latest versions are manufactured in India, even though defence spending under Nehru was low enough to account for India's reversals in border conflicts.

China has long replaced India as a model for developing countries in western thinking, given all its anti-Soviet rhetoric, and its close defence and economic ties with the United States.

Unfortunately, democratic institutions in India provide enough dissent to verge on myth and propaganda. You will nowhere experience this sense of freedom and individual dignity, supposed to be the essence of western values, in a China bent upon change along occidental lines.

M. K. AGARWAL

Laboratoire de

Physio-Hormono-Réceptérologie,
(Université de Pierre et Marie Curie),
15 Rue de l'École de Médecine,
75270 Paris Cedex 06, France

SIR—N.H. Antia's letter criticizing India's scientific progress (*Nature* **331**, 384; 1988) is unduly harsh and emotional. If the "capitalist West" has a "morbid fear of communism", then Antia's letter reveals a distorted view of India and adoration for so-called communism. Many of the problems of the poor in India are due to the increasing population, which has nearly doubled since independence, and rises more rapidly in economically lower strata of society. China has been able to arrest the sharp rise in population at gunpoint. In China, it is not the individuals who decide how many children they should have but the state.

India's recent history shows that the suppression of freedom by Mrs Gandhi pushed her out of office. The result of an election does not depend just on the votes of the 'privileged' but on those of the masses, which shows how important freedom is for an average Indian. It would

have been shameful for India to take the path of China in order to make more rapid scientific, technological and material progress and to pay the price not only with the suppression of freedom of expression and movement, but also by sacrificing the most vibrant and oldest surviving traditions in the world.

Indian scientists and doctors going abroad speak for India's great scientific and medical awareness and competence in global participation in those two fields. The rural population moving to urban areas and living there in bad conditions is a transitional phenomenon of any industrial revolution, as history shows. "Freedom for a few only" are the words used by communists or latent communists even in the affluent West.

Antia should take a balanced view of the problem and take note of the price China has paid in terms of its culture and tradition as well as human lives and suppression.

It is better for us, as scientists, to take a more positive view in order to cure the evils of Indian society rather than condemning its achievements in science and technology. Those who condemn are also the first to be outraged at the suppression of any freedom. It is easier to criticize and condemn a system when one is allowed to do so. If such people were forbidden to travel from Bombay to Poona without the permission of the authorities, they would realize the value of freedom.

V. VERMA

I. Physiologisches Institut der
Universität des Saarlandes,
D-6650 Homburg/Saar, FRG

Unjust Congress

SIR—We were distressed by your report (*Nature* **332**, 670; 1988) on the recent congressional hearings regarding fraud in science. Your article merely repeated the various allegations made at the hearings by Drs Margot O'Toole, Charles Maplethorpe, Ned Feder and Mr Walter Stewart regarding the paper by Weaver *et al.* that appeared in *Cell*.

As the three scientists who, on O'Toole's request, reviewed the data on which the *Cell* article was based, we feel that other views should have been aired, not just the charges. Your failure to do this perpetuates the injustice generated by hearings in which none of the scientists who performed the relevant experiments or participated in the reviews was asked to testify. The result is that a one-sided version of events has been put before the public.

O'Toole initially turned to us as friends to seek our help and judgement on what to her seemed evidence of fraud involving the article in *Cell*. Her accusations were not based on her own work at Massachusetts Institute of Technology (MIT), but on some notebook data that she had

come across by chance. After reviewing the data and consulting the involved parties, we unanimously concluded that there was (1) no sign of fraud; (2) no evidence of misrepresentation; (3) no error that undermined the article's basic conclusion. Contrary to O'Toole's statement at the hearings, we did not concede that her criticism was sound.

It was suggested at the hearings that the whistle-blowers in this case have sacrificed their careers by questioning the science of senior investigators. To our knowledge, nothing was done to impede O'Toole in making an official complaint to MIT or *Cell*. On the contrary, she testified that she was encouraged to ask for an official inquiry but chose not to do so. We are not aware of steps that she has taken to continue her career, nor have we, or anyone to our knowledge, made any attempt to block her in this endeavour. Furthermore, the other person who raised charges of fraud, Dr Charles Maplethorpe, is still in science.

Up to the present, the scientific issues have not been put before the public. We thus welcome the independent scientific investigation being organized by the National Institutes of Health. But a picture depicting the authors of the *Cell* article as guilty has been created, and we fear that no matter what results from the official inquiry, an after-image will remain.

It has always been our belief that the most important test of a scientific claim is independent experimental verification, not judicial review. We hope that the editors and readers of *Nature* share this view.

HENRY H. WORTIS

BRIGITTE T. HUBER

Department of Pathology,
Tufts University School of Medicine,
136 Harrison Avenue,
Boston, Massachusetts 02111, USA

ROBERT T. WOODLAND

Department of Molecular Genetics
& Microbiology,
University of Massachusetts
Medical School,
55 Lake Drive North,
Worcester, Massachusetts 01605, USA

Life begins at . . .

SIR—In their paper on human gene expression¹, Braude *et al.* use the term "pre-embryo", though obviously with a certain reserve as they were careful to reference the source². The term itself is not an objective, well-defined scientific descriptive, but in its origins and application it is a mere administrative device to obviate the legal and ethical considerations limiting experiment on human entities at more advanced stages of development, however far that ulterior motive may be from the intentions of the authors.

Indeed, the experimental results reported by Braude *et al.* reveal the lack of a scientific basis for the prefix, given that they have established a specific biochemical effect characteristic of human individuality already at the 4- or 8-cell stage, in the expression of its distinctive genes.

I would therefore expect that this example of subjective and arbitrary terminology be carefully excluded from the scientific literature, the considerations of Chargaff in a recent Commentary in *Nature* being surely of some relevance to this issue³.

A.J. McEvoy

Department of Chemistry,
Institute of Physical Chemistry,
EPFL-Ecublens,
CH-1015 Lausanne,
Switzerland

1. *Nature* 332, 459 (1988).

2. 1st and 2nd Reports of the Voluntary Licensing Authority for Human Fertilisation and Embryology (Medical Research Council, London, 1985 and 1986.)

3. *Nature* 327, 199 (1987).

PhD theses

SIR—Lars H. Breimer (*Nature* 332, 481; 1988) paints too rosy and chauvinistic a picture of the Swedish PhD system, which has some disadvantages. It is true that a Swedish thesis is often based on four or five articles published in journals but this is not a statutory requirement. In fact, the Swedish statutes allow as equally valid alternatives a collection of articles (single or multiauthored) not published in this way as well as a monograph-type, previously unpublished thesis.

Because a thesis in Sweden has been given an ISBN number, printed and distributed before its public defence, it remains registered as a PhD thesis in the university library, whether or not it is passed by the examiners' committee. There is no provision in the Swedish system (as there is in the British) for revision and resubmission of a thesis, a major drawback that sometimes amounts to pressure on the examiners to give a borderline candidate the benefit of the doubt rather than fail him/her. And according to the Swedish Universities Statutes, neither the reasons for the acceptance of a thesis nor any dissent in the examiners' committee meeting may be reported in the minutes of the meeting or in any other document (Chapter 8, Article 37, paragraph 5). The majority decision of the committee is final.

The printing and mailing costs for the statutory number of copies are paid by the faculty only up to a certain maximum amount. Local faculty rules may require that 100–150 copies be made available before the defence. Although the costs are paid in part, the work of addressing, mailing and delivering copies is done largely by the secretarial staff of the department concerned and its cost is not

reimbursed. Even if the thesis consists of published journal articles, these must also be supplied in the required number. The system tends to be wasteful of work and material.

The public oral defence is potentially a valuable procedure that should be retained but a candidate is rarely failed once the thesis has reached this stage. Such an event creates newspaper headlines. I can recall only two cases of rejection in Sweden in the past 15 years.

The upshot is that although the Swedish system has many good points, it is in need of some overhaul.

As to the part of Breimer's letter concerning the situation up to about 25 years ago, it is true that there used to be three opponents. One of them was, however, nominated by the candidate himself and confined himself largely to pointing out undotted i's, uncrossed t's, punctuation errors and the like. The third opponent, also nominated by the candidate, was invariably a ceremonial figure, who made witty remarks at the end about the thesis (no real criticism). Even in those days, a thesis could almost never be failed at this stage although one redeeming feature of the old system used to be that the thesis could be given a grade (1 to 5 or rather on a letter scale C to A) instead of only passed or failed as in the present system.

D.S. PARASNIS

Department of Applied Geophysics,
Luleå University of Technology,
S-951 87 Luleå,
Sweden

SIR—In "The thesis that won't go away" (*Nature* 331, 497; 1988), Beverley Halstead puts forward a number of important but contradictory views. As research students, we would support the suggestion that a PhD should be a period of apprenticeship but we do not agree with the idea that the only way to assess the worth of a scientist is by measuring the volume of work published. In order to demonstrate one's competence as a research scientist, is it absolutely necessary to reach a point at which the work is suitable for publication? Indeed, such a requirement could feasibly lead to a situation in which the integrity of the work is sacrificed for prompt and plentiful publication.

In addition, the successful completion of a project is not solely dependent on a student's ability as there is great variation between PhD projects with regard to difficulty, supervision and availability of resources. A student in a well-funded laboratory who is part of a large group may receive greater stimulation and help than a student of equal ability struggling on his or her own in an ill-equipped laboratory. Furthermore, a student who focuses on a single problem with the aim of publishing the data may become a less competent research scientist than one who

has been encouraged to take a more holistic view of his or her work. It is already apparent that the pressure to produce a thesis encourages students to consider only those areas within their own field which are of immediate relevance to their project and the pressure to publish inevitably increases the risk of their pursuing their subject narrow-mindedly.

The thesis system is certainly not perfect, but it is still the fairest method of establishing whether or not a student is worthy of a doctorate. We believe there is urgent need for change within this system rather than its replacement with another. Most importantly, an attempt must be made to unify across universities the standards of assessment used by individual examiners.

ANTHONY P. HOLLANDER

BUBBLY DULARAY

Department of Pathology,
University of Bristol,
Medical School,
University Walk,
Bristol BS8 1TD, UK

SIR—A. J. Greenfield's letter (*Nature* 332, 481; 1988) about the apparent disparity in the remuneration of PhD students and PhD research assistants is misleading.

The salient differences, which he omits to mention, may be summarized as follows. The PhD student is generally a new graduate with few financial responsibilities, and can therefore afford to take a higher degree which will enable him or her to pursue research in any field which takes his or her fancy.

A PhD research assistantship, however, is preferentially given to someone with a number of years' relevant experience in industry or elsewhere on top of a good first degree, and is therefore likely to be considerably older with many more financial commitments. His salary is being paid by an institution which requires a specific piece of research to be undertaken, and about whose direction the researcher has very little say. He is therefore being paid to do a job of work, after which he will no doubt be required to produce a report of his findings to his sponsor.

The fact that a revised copy of the report may be submitted as a thesis and offered to the research institution for evaluation is irrelevant. The PhD is a bonus which is rightfully gained for having satisfied the academic criteria governing such research.

The saving grace of this alarming debate is that, given the historic traditions of the better universities, it will be a very long time indeed before any change is seen in the present, and eminently satisfactory, system.

S. J. SHEPHERD

Corner Cottage,
St John, Torpoint,
Cornwall PL11 3AW, UK

Reticence and the upper limit

People who have nothing much to say, but that they have failed to substantiate another other fellow's claims, seem to follow those with great boasts to make by concealing their messages.

THIS seems to be the era of the phenomena which are not, of the startling observations which seem for a time to stand the world on its head, but which are not afterwards confirmed. By now, almost everybody is familiar with the research reports whose abstracts finish with some such sentence as "... we have therefore established an upper limit of 2×10^7 " with just the suspicion of being the announcement of a discovery in its own right and not the end of some other poor fellow's hopes.

For those with a taste for the macabre, a recent (2 May) issue of *Physical Review Letters* seems to have an unusually rich harvest of upper-limit articles. There is, for example, another contribution to the young but now-rich literature of the fifth force (V. L. Fitch, M. V. Isaila and M. A. Palmer, all Princeton physicists).

The case, it will be recalled, is that of the announcement by E. Fischbach *et al.* at the beginning of 1986 that inconsistencies in the original data of Eötvös on the gravitational attraction between objects of different materials could be neatly explained if there is an "extra" force of intermediate range dependent on the baryon number (how many nucleons) of the interacting materials.

It is admirable in the true Latin sense, but not really surprising, that the rich crop of experimental articles since to have appeared are nicely contradictory. At a guess, perhaps one article in four puts flesh on the Fischbach conjecture, which was experimental only in the sense of being a re-examination of data collected more than half a century earlier. The others are "upper-limit" articles which, however objective, tend to reinforce the impression that the fifth force may have made only a brief appearance on the scene (not that that will deter people from trying to measure the force, nor should it).

Fitch *et al.* have shown what remarkable steps people will take to secure data in the field. They have repeated the Eötvös measurements on the north-sloping face of Mt Maurice at Yellowstone, Montana. A hint of the difficulty is provided by the tangent of the slope of the mountainside to the South (and above), quoted as 0.5.

Yet, fair play, there is not a hint of disappointment at the obvious scatter of the crude measurements, or in the estimates of the parameters defining the supposed extra force which are all accompanied by errors larger than themselves.

Laconically, the abstract says that "...we find the product of strength a_0 and range r to be -0.04 ± 0.07 m...". In the text, the authors say "...these results agree with those in Refs. 8, 9 and 11, and are in disagreement with those in Ref. 12..." without telling their readers that "Ref. 12" is P. Thieberger's affirmative measurement of the fifth force on the Palisades cliffside opposite Manhattan. Those wishing to read upper-limit articles intelligently must first learn the sign-language of the calculated understatement.

Some of the same applies to a more spectacular upper-limit article in the same issue by a group (B.L. Dingus *et al.* 60, 1785; 1988) scattered between the University of Maryland, Los Alamos, the University of New Mexico and the University of California, Irvine, which has been looking for evidence that Cygnus X-3 (the X-ray star with a binary orbital period of 4.8 hours) is a powerful source of energetic γ -rays, with energy exceeding 10^{16} eV. Plainly it is a matter of some importance that the issue should be pinned down.

The need persists for sign-language in reading this article as well, but one is helped along a little more. The inference that the periodically emitted energetic particles must be γ -rays derives from a simple estimate that only electrically neutral particles could successfully navigate intergalactic space without dispersion that would entirely obscure the 4.8 hour periodicity observed in the X-ray flux, but there have been (since 1983) reports of μ -mesons as well. So Dingus *et al.* have built a powerful cosmic-ray shower detector at Los Alamos, giving themselves the luxury of detecting accompanying muons of sufficient energy (above 2 GeV) individually.

Sadly, the outcome of all this sophistication is that the periodicity of Cygnus X-3 as a source of 50 TeV γ -rays is not confirmed — but you would not guess that from the abstract. The skeleton of the main sentence is "Upper limits for ... the flux of air-showers from ... Cygnus X-3 ... are found to be 1×10^{-13} cm⁻²s⁻¹ for all events, and ... for all events with no observed muons...". If one asks how these fluxes compare with those which have stimulated these measurements, one must presumably go burrowing in the references.

But this paper does have two saving graces. First, buried away towards the end of the text is an explicit sentence reading "data from the entire run show no statistically significant signal ... from the direc-

tion of Cygnus X-3." Second, there follows simple declaration that there was a period of 45 days (out of a total of 264) when there were 27 extensive showers that might have been correlated with the phase of the distant X-ray star even though "... the statistical evidence is not compelling".

That, for what it is worth, is not the end of it. The same issue of the same journal has no fewer than two calculations of upper (or lower) bounds for the properties of the particle called the axion derived from the observation of neutrinos from last year's supernova, 1987A. (One of them, without as much inconsistency as there might seem to be, derives both an upper and a lower bound for the axion mass, but puts the former below the latter: the argument is that calculations of what will happen to axions depend crucially on whether they are massive enough to be trapped in the supernova gravitationally, or whether they move without constraint.)

There is a temptation to make fun of such logic-chopping, but there is a practical problem to be solved by those who engage in it, or on their behalf. First, as with the fifth force, it is in no sense surprising, but the opposite, that the announcement of a new phenomenon should be followed by a rash of confirmatory and contradictory results. By definition, the first measurement will have been at the edge of what is technically possible; those that follow, at least at first, are likely to be less expert and therefore less sensitive. That is why those breaking new ground should not be downcast that most of those who flatter them by repeating their measurement end up in disagreement.

Second, the credentials of those who repeat others people's measurements should be examined, not so much for evidence of cynicism but the opposite. There may be occasions when a person sets out to repeat another's experiment in the expectation that he will find an error, but more often it must be the case that the persons who spent all those months building equipment must have believed they would find something, not find something else wrong. Third, there is a matter of prose style. People who may be prevented, by modesty or some such consideration, from saying in plain language what they have found out, might be expected to name names when their conclusion fails to confirm what somebody else has claimed. But the convention seems to go the other way, which is a pity.

John Maddox

Polymer science

Conducting rubber bounces back

Paul Calvert

THE conventional picture of conducting polymers, as typified by polyacetylene ($-\text{CH}=\text{CH}-$)_n, is that they are black, rigid, conjugated (having alternating single and double bonds) and insoluble. Their properties have a closer resemblance to coal than to polyethylene. M. Thakur of AT&T Bell Laboratories now makes the surprising claim (*Macromolecules* **21**, 661–664; 1988) that conjugation is not necessary for a polymer to be conducting, and that natural rubber (*cis*-polyisoprene) becomes conducting when treated with iodine. This raises the disconcerting possibility that the rest of us may have been overlooking the obvious.

The *trans* form of polyacetylene has a conductivity of about 10^{-5} siemens (S) cm^{-1} , which makes it a semiconductor. The polymer is black and shiny because the conjugation, the extended sequence of alternating double and single bonds, reduces the energy gap between the fixed, valence-electron states and the free, conduction states: visible light is absorbed by electrons jumping from the valence band into the conduction band. Reduction of the polymer with alkali metals puts electrons into the conduction band and makes the polymer more conducting. Similarly, reaction of the polymer with an oxidizing agent, such as iodine, forms a charge-transfer complex. This removes a valence electron, leaving a hole in the valence band which can also travel through the polymer and conduct. Normal iodine-doped polyacetylene can reach conductivities above 100 S cm^{-1} .

At first it was thought that the conduction arises from electrons or holes travelling along chains, and then leaping the gaps at the ends. It later became clear that conductivity is dominated by carriers jumping from chain to chain, because the paths along the chains are normally blocked by irregularities. This was proved last year by the formation of polyacetylene with very high conductivities, in which the polymer chains have been fully straightened (Basescu, N. *et al.* *Nature* **327**, 403–405; 1987 and my accompanying News and Views article on page 371 of that issue).

Conjugation was nonetheless thought to be necessary because the electrons have to be delocalized over a large distance to produce a small energy gap between the valence and conduction bands. This extended orbital structure also locks the chain into a rigid conformation such that it cannot melt or dissolve. Polyisoprene ($-\text{C}(\text{CH}_3)=\text{CH}-\text{CH}_2-\text{CH}_2-$) has an isolated double bond with no conjugation. Every chemist knows that rubber is

attacked by bromine at the double bond, and that rubber bungs in iodine bottles go black. Thakur measured conductivities, after iodine doping, of 10^{-2} – $10^{-1} \text{ S cm}^{-1}$, 10 orders of magnitude above normal rubber. He believes that the iodine forms a charge-transfer complex and has some infrared evidence to support this.

There are two worrying aspects of this work. First, it makes us all look stupid. There have been studies of various composites of polyacetylene and rubber, but apparently no-one noticed doping and conductivity in the rubber component. Second, the effect is seen with *cis*- and *trans*-polyisoprene (natural rubber and gutta-percha, respectively), and also with poly-2,3-dimethylbutadiene ($-\text{C}(\text{CH}_3)=\text{C}(\text{CH}_3)-\text{CH}_2-\text{CH}_2-$) but not, however, with poly-1,4-*cis*-butadiene ($-\text{CH}_2-\text{CH}=\text{CH}-\text{CH}_2-$). Thakur suggests that this is because of the methyl

(CH_3) groups increasing the electron density in the double bonds, but this effect is not normally large.

Polyphenylacetylene attracted attention as a conducting polymer some years ago, but it was shown that the effect arises from ionic conductivity by I^- and not electronic conduction (Cukor, P. *et al.* *Makromolekul. Chem.* **182**, 165–171; 1981). Thakur checked this possibility by long-term direct-current measurement where ions would be expected to become depleted so that the conductivity drops off. No drop-off with time was seen. Polyphenylene-sulphide becomes conducting when doped with arsenic pentafluoride, because of a chemical reaction that forms a new, dopable polymer (Frommer, J.E. *et al.* *Am. chem. Soc. Symp.* **242**, 447–453; 1985). Thakur's spectroscopic studies show no evidence of such a structural rearrangement. Thus one's prejudice says that there must be a catch, but this report by Thakur is a carefully constructed piece of work with no obvious loopholes. If it does stand up, we shall have a lot of rethinking to do. □

Paul Calvert is in the School of Chemistry, University of Sussex, Brighton BN1 9QF, UK.

Particle accelerators

The light that never was

Roger G. Evans

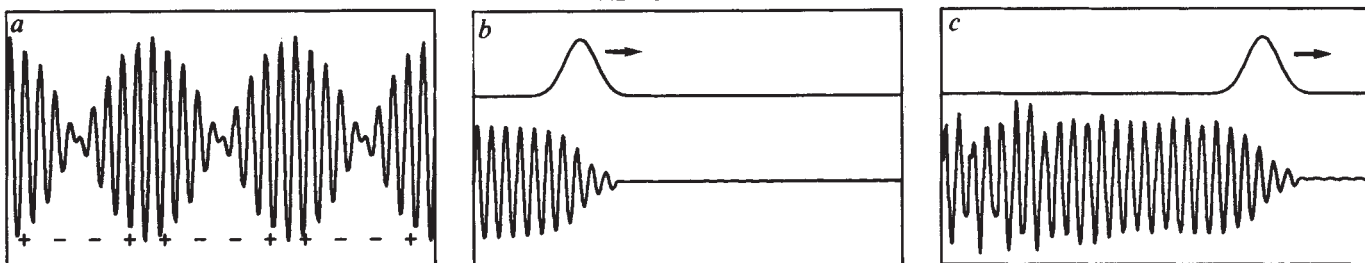
As physics has progressed, the secrets of 'elementary' particles and their interactions have successively become more difficult to unlock, and are tied up in smaller and smaller objects—from atoms to nuclei and now to quarks. The Heisenberg uncertainty principle requires that smaller distances can only be probed by higher and higher energies and has led to the building of immense particle accelerators, for instance at CERN and in the US national laboratories. Probing smaller objects also means that more particles per unit area are required to give a reasonable interaction rate, and the need for a high centre-of-mass energy leads to concentrating on head-on colliding-beams systems. H. Hora proposes elsewhere in this issue (*Nature* **333**, 337–338; 1988) an unconventional accelerator based on intense laser beams that can raise particle energies to 10^{12} electron-volts (1 TeV) without being unmanageably large. It shares, however, many difficulties of other accelerator designs using nonlinear effects.

Over the past half century, the designers of particle accelerators have a proud record of doubling the particle energy every few years. This progress has been possible only by continual innovation, and since about 1980 there has been a concerted effort to find the successor to the conventional radiofrequency linear

accelerator. For a while, the combination of lasers to generate exceedingly high fields, and ionized plasmas to provide a resonant medium with a longitudinal electric field, seemed to promise a compact design for new machines. But plasmas are bedevilled with instabilities which could ruin the beam quality of an accelerator. Not surprisingly we now have, in Hora's proposal, an imaginative attempt to design a laser-based accelerator without a plasma.

Hora postulates a phase-modulated light wave, in which a nonlinear (or ponderomotive) force is somewhat enhanced and electrons initially in the laser beam are expelled with a velocity rather greater than their initial oscillation velocity. The fact that the required light wave fails to satisfy Maxwell's equations of electromagnetism is not the main objection to the application of this scheme. Rather, the device is analogous to electrostatic accelerators such as the van der Graaf, and it is not clear how stages can be cascaded together to achieve the desired particle energies.

Having to live within the rigours of Maxwell's equations is a long-standing problem in attempts to use lasers to accelerate charged particles. In a vacuum, the electric field of an electromagnetic wave is transverse and does not provide



Laser-beat accelerator for electrons. *a*, The beat of two laser waves pushes electrons into the regions of minimum light intensity; *b* and *c*, as the laser pulse (upper line) propagates, it leaves behind a 'wake' of intense plasma waves.

acceleration forwards in the manner of a conventional radiofrequency cavity. To provide a longitudinal component of field it is necessary to introduce a material medium (such as plasma) or a material boundary within a few wavelengths of the beam (the laser-grating accelerator). Solid or liquid materials always suffer from some limiting field strength of electrical breakdown and limit the rate of acceleration that is achievable. Plasmas, on the other hand, have already broken down, so that the electric field is limited only by plasma density and can be a factor of 100 or more greater than for conventional means (say 10 GeV m^{-1} compared with 100 MeV m^{-1}), with a corresponding reduction in the size of the accelerator.

Probably the most extensively studied laser-plasma accelerator is the beat-wave system based on two parallel laser beams of slightly different frequencies in a plasma resonant at the difference (or beat) frequency (see figure). The plasma acts as the resonant cavity of a radiofrequency accelerator and enables the wave to be pumped over a relatively long period, easing the demands on laser power and giving an accelerating field which can comfortably exceed the electric field of the laser.

A high electric field alone, although it makes for a compact accelerator, is not the only figure of merit. The requirement to focus the particles at the end of the accelerator is unbelievably stringent. Current designs for a $1 + 1 \text{ TeV}$ electron-positron collider would need to focus their beams to about $100\text{-}\text{\AA}$ spots to achieve a usable luminosity. Plasmas would be subject to instabilities such as stimulated Raman scattering caused by the laser, or Weibel instabilities caused by the beam particles, which could prevent the beam being focused.

Methods have been proposed to reduce plasma instabilities by reducing the laser pulse length, but these would make the laser more difficult and expensive to construct, and herein lies perhaps the main problem — lasers are a very expensive power source to drive an accelerator. The hypothetical 1-TeV machine mentioned above could require a power consumption approaching one gigawatt (requiring a medium-sized power station).

Although it still suffers the economic limitations of lasers, the scheme proposed

in this issue by Hora is undoubtedly ingenious. The acceleration of electrons in a non-uniform light wave is a well-known phenomenon, Hora himself having published many papers on the topic. Essentially, the oscillatory motion of the electron in the vibrating electric field of the light is gradually converted to linear motion by the non-uniformity of the electromagnetic wave. The time average of the high-frequency forces felt by the particle — the ponderomotive force — is derivable as the gradient of a ponderomotive potential which is proportional to the light intensity divided by its frequency.

The novel feature of Hora's scheme is that by phase modulating the laser, he introduces low-frequency components which enhance the ponderomotive force. Hora explains how the phase modulation could be produced by a lens with an intensity-dependent refractive index, but does not calculate in detail how the intensity at the lens surface is related to the intensity in the accelerating region. It seems likely that problems with laser damage to the lens would be severe. Attempts to phase modulate at low intensity would be frustrated by the intended

mode not being a true solution of the vacuum wave equations, and not propagating successfully. Together with the problems in cascading separate accelerating stages, this makes the paper of more academic than serious practical interest.

The light intensities referred to by Hora seem very large by the standards of current experiments but they are not unreasonable extrapolations of current work on amplifying sub-picosecond laser pulses to high energies. It is intriguing to extrapolate just a little further, to intensities of about $10^{23} \text{ W cm}^{-2}$ for which the oscillatory energy of electrons in the laser beam would be about 1 TeV . Electrons and positrons would of course oscillate in antiphase in such a field and one immediately has a collider. At these intensities there is no need for a plasma, or for any departure from Maxwell's equations. Perhaps the attempts to harness lasers into high-energy physics were a little premature and practical developments over the next few years will bring even more imaginative applications. □

Roger G. Evans is Head of the Advanced Research Computing Unit, Rutherford Appleton Laboratory, Didcot OX11 0QX, UK.

Recombinant DNA

Dos and don'ts of genetic release

Stuart Glover

IT IS 14 years since the recombinant DNA debate began. Recombinant DNA experiments conducted under the guidelines and regulations established in the United States, in Europe and elsewhere have been carried out on an ever-increasing scale since then without harm to the health of people engaged in this technology. As a result, regulations have been lessened in rigour as experience has accumulated. The debate has moved from the laboratory to the field now that the intentional release of genetically engineered organisms into the environment is being sanctioned. The scientific issues relating to the release of genetically engineered microorganisms were discussed at a recent meeting*.

There have been very few examples of the release of genetically engineered

microorganisms into the environment, so we have little knowledge of their ability to survive and persist. But quantities of bacteria are released into the environment continuously in sewage; and millions of hectares of land are inoculated with *Rhizobium* each year to improve the growth of leguminous crops, providing good model systems. One important lesson from such studies (J.E. Beringer, University of Bristol) is that introduced strains will not usually compete successfully for nodule formation where indigenous rhizobia are well established. Over the past few years, rhizobia carrying transposable genetic elements conferring drug resistance have been released to field soils to study the fate of the genes carried in plasmids.

Methods for the tracking and monitoring of released genetically engineered

* Release of genetically engineered microorganisms (REGEM-1), Cardiff, 5-8 April 1988.

microorganisms have improved markedly in recent years. The *lacZY* genes of *Escherichia coli* delivered to fluorescent pseudomonads by an engineered Tn7-*lac* element are a highly effective tracker-gene system and were released in a test approved by the environment protection agency (EPA) in the United States last year (G.F. Barry, Monsanto, St Louis). Both vertical and horizontal movements of the pseudomonads away from the site of inoculation were monitored and found to be negligible.

Other methods include the use of immunofluorescent techniques based on either polyclonal or monoclonal antibodies to detect specific organisms. The detection limit of bacteria (*Thiobacillus ferrooxidans*) is 10^6 – 10^7 cells per gram of dry sediment by immunofluorescence or 10^5 – 10^6 cells per ml by enzyme immunoassay (G. Muyzer, University of Leiden). Nucleic-acid hybridization techniques based on 5S or 16S RNA sequence databases provide another even more sensitive detection method (D. Stahl, University of Illinois). There is much promise (R.M. Atlas and R.J. Steffan, University of Louisville) of methods based on target DNA amplification using the polymerase chain reaction (PCR), an *in vitro* method for increasing the number of copies of a target nucleotide sequence. By this method, target sequences can be amplified by a factor of more than 10^7 in a few hours. Coupling PCR with direct extraction of DNA permits the detection of *Pseudomonas cepacia* at concentrations as low as 1 cell per gram in sediments containing 10^9 diverse non-targeted bacteria per gram.

Risk assessment

Detailed risk assessment and environmental impact analyses preceded the approved field trials of genetically engineered baculoviruses as insecticides. These studies demonstrate that an innocuous genetic marker is a suitable tool for the identification of an engineered virus released into the environment. They also show that a genetically crippled virus can be constructed that degrades in the environment after its effect on a caterpillar host. Such a virus is a suitable subject for further engineering, and experiments are under way to construct viruses that act more quickly than the natural virus through the incorporation of genes for insect specific toxins and hormones (D. Bishop, NERC Institute of Virology, Oxford).

Equally rigorous risk assessment and environmental impact studies preceded the approved field trials of recombinant ice-minus strains of *Pseudomonas syringae* for biological frost-damage control of potatoes (S. Lindow, University of California). The consequences of the release have been carefully monitored and reveal that ice-minus *P. syringae* were not

detected at any time on vegetation surrounding the experimental plot; no ice-minus bacteria were detected at any time in surface water surrounding the plot; ice-minus strains did not survive well in the soil and did not succeed in colonizing vegetation surrounding the plot (R. Seidler, EPA Corvallis). Five different types of sampling devices were used: (1) six-stage Andersen samplers; (2) Reynier slit samplers; (3) all-glass impinger samplers; (4) gravity plates; and (5) sentinel potted plants. These studies exemplify the success that can be achieved when experimenters and regulatory agencies understand one another's problems.

EPA approval for the field testing of *Pseudomonas fluorescens* engineered to express a non-pesticidal trait (lactose use) has been granted as a prelude to the evaluation of introducing *Pseudomonas* root colonizers containing the *Bacillus thuringiensis* δ -endotoxin gene as an insecticide. The release experiments have not yet been conducted, but field experiments with tomato plants containing the δ -endotoxin gene leave no doubt about the efficacy of the toxin gene. Untreated plants were severely damaged under conditions where recombinant plants containing the toxin gene were not (L. Watrud, Monsanto, St Louis).

Since the development of vaccinia virus as an expression vector, a wide range of foreign DNA sequences which encode antigenic determinants of bacterial, viral and parasitic origin have been inserted at non-essential regions within the viral genome. Inoculation of such recombinant viruses into appropriate species elicits an immune response which in most cases is protective against challenge by the live pathogen. In a novel departure, fowl pox virus (FPV), which has obvious applications for use as a vehicle for vaccination of poultry, has been developed for use in non-avian species (E. Paoletti, Virogenetica Inc., Albany). Non-permissive (non-avian) cell lines do not allow replication of the virus, which nevertheless appears to persist and in these cell lines recombinant viruses express inserted genes.

Recombinant FPV containing the rabies virus glycoprotein gene has been tested in seven animal species (chicken, mouse, rat, rabbit, cat, dog and cattle), all of which responded by the production of anti-rabies antibody. Two dogs and two cats subsequently challenged with rabies virus survived the challenge and remained healthy and without symptoms. All non-vaccinated animals similarly challenged died. The potential use of FPV recombinants as vaccine vectors for non-avian species has some important advantages. In particular, the non-productive aspect of the vector adds an additional safety feature into the procedure as the

possibility of transmission of the recombinant to non-vaccinated individuals and to other species is greatly reduced or eliminated.

Recombinant vaccinia-Sindbis virus has been used as a model for studying the efficacy of live viral vaccines in calves. The experiment carried out in New Zealand with approval of the appropriate authorities demonstrate that the virus was not transmitted to control animals penned with inoculated cattle. Antibody against both the recombinant virus and the Sindbis virus was demonstrated (E. Wedman, EPA Corvallis). These inoculated animals were not challenged with live Sindbis virus so that the effectiveness of the vaccination remains to be established.

Unauthorized use

On the other hand, no approval from appropriate authorities in Argentina was obtained for a field trial in Azul conducted by investigators from the Wistar Institute (Philadelphia) and the Pan American Health Organization in September 1986 (see the news story by Joseph Palca in *Nature* 324, 202; 1986). The experiment involved the testing of a vaccinia-rabies recombinant vaccine in cattle. It was not so much the experiment itself that was the subject of criticism, but the way it was planned and implemented. Preliminary results of monitoring since the experiment, now interrupted by the Argentinian authorities, show that all vaccinated and contact animals were seroconverted 6 months after the start of the experiment. Analysis of blood samples from three of the four local animal caretakers indicate that one was seroconverted for rabies virus antibodies.

The recombinant virus appears therefore to have passed from vaccinated animals to all of their animal contacts and to humans involved in handling and milking these animals. Other possible animal contacts, such as calves and rodents, are now being tested to determine the extent of the spread of the virus (J. La Torre, Centro de Virologica Animal Serrano, Argentina).

The scientific issues relating to the release of genetically engineered microorganisms are becoming much clearer where they relate to engineering organisms for safety and for specific use. Methods for detecting and monitoring their survival and persistence are being improved and methods for measuring their behaviour and functioning in model ecosystems are now available. Although the number of approved release experiments is small, the lessons learned from the careful monitoring that has followed will be useful for future regulation. □

Stuart Glover is Professor of Genetics at the University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.

Evolutionary biology

Genetics and labour in bees

Michael D. Breed

EARLY models for the evolution of worker caste in social insects assumed that there is little genetic variation in the social unit¹. But recent genetic studies show that colonies of many species incorporate considerable variation², either because many queens are present (polygyny)³ or because the queen has mated more than once (polyandry)⁴. In this issue, Robinson and Page⁵ on page 356 and Frumhoff and Baker⁶ on page 358 demonstrate that genetic variation within bee colonies is correlated with differences in task performance. This indicates that there is a genetic basis for behavioural differences among workers, and the result has substantial implications for our understanding of the evolution of division of labour in colonies of social insects.

The two papers in this issue suggest that polyandry in the honey bee broadens the range of abilities that workers in a colony have for performing certain tasks. Robinson and Page⁵ study two behavioural patterns, guarding and undertaking (removal of corpses from the hive; see this week's cover), that are critical to colony function but are expressed by very few bees in the colony. They can distinguish worker patrilines — worker offspring of different fathers but the same mother — by using electrophoretically detectable enzyme markers. Certain patrilines show a strong bias for guarding, whereas others show a strong bias for undertaking.

Frumhoff and Baker⁶ look at two behavioural patterns in which workers commonly engage during social interactions: grooming of workers by other workers and mutual feeding (trophallaxis). Cuticle-colour genetic markers (black and cordovan) distinguish patrilines in their colonies. Cordovan patrilines engage in substantially more grooming behaviour than black patrilines. There is no consistent bias in feeding behaviour; Frumhoff and Baker suggest that this difference results from the fact that some workers specialize in grooming whereas all workers engage in trophallaxis.

Crozier and Page suggested⁷, among other hypotheses for the evolution of polyandry in social insects, that a colony with high genetic variability among its workers might cope better with changing environmental conditions. Cole argues⁸, however, that polyandry is associated with large colony size, and that polyandry ensures adequate sperm supply for the production of large populations. Recently, Sherman *et al.*⁹ have proposed that the increased genetic variation in colonies resulting from polyandry might buffer the

colonies from the effects of parasites and pathogens.

The findings reported in this issue^{5,6} suggest that the high number of matings, which allow the queen to obtain sperm far in excess of the amount that can be stored or used, may have evolved as a result of selection to produce behavioural genetic variability in the worker population. But both groups of authors are careful to point out that they may have presented epiphenomena, secondary to other factors which might have led to polyandry in the honey bee.

Are queens able to select the drones they mate with to maximize the range of worker inclinations in their colony? Each of the studies in this issue used colonies whose queens were artificially inseminated from limited numbers of drones. In nature, queens mate up to 17 times¹⁰, so one next step would be to determine if these matings are non-random; that is, whether queens seek to mate with a varied sample of males.

Polygyny in social insects, particularly in ants, is common; electrophoretic studies of polygynous ants reveal low levels of relatedness (and consequently high genetic variation) among workers in such colonies³. Polyandry is less common, but is more difficult to demonstrate than

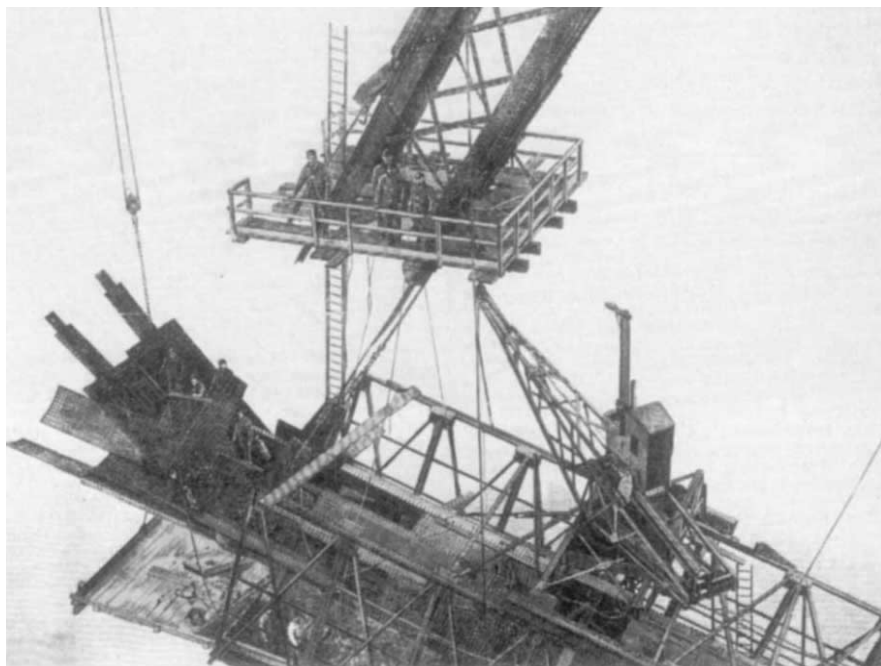
polygyny. It will be interesting if comparative studies on polygynous species reveal the same correlation between genetic background and proclivity to perform a task.

Polygyny and polyandry are usually assumed to have evolved secondarily to the presence of a worker caste. Both these mechanisms reduce genetic relatedness within colonies. Genetic models of social evolution predict that reduced relatedness should result in increased conflict of interest among workers and selection should favour reproductive capability in the workers, which are normally nonreproductive. Results such as these, which indicate continued social cooperation even as genetic variation has increased, suggest either that workers evolved in such a way that they are locked into cooperative behaviour patterns, or that we must consider models of oppression for the evolution of worker sterility¹¹. □

1. Hamilton, W.D. *J. theor. Biol.* **7**, 1–52 (1964).
2. Pamilo, P. *Heredity* **48**, 95–106 (1982).
3. Holldobler, B. & Wilson, E.O. *Naturwissenschaften* **64**, 8–15 (1977).
4. Page, R.E. & Metcalf, R.A. *Am. Nat.* **119**, 263–281 (1982).
5. Robinson, G.E. & Page, R.E. Jr *Nature* **333**, 356–358 (1988).
6. Frumhoff, P.C. & Baker, J. *Nature* **333**, 358–361 (1988).
7. Crozier, R.H. & Page, R.E. *Behav. Ecol. Sociobiol.* **18**, 105–115 (1985).
8. Cole, B.J. *Behav. Ecol. Sociobiol.* **12**, 191–201 (1983).
9. Sherman, P.W., Seeley, T.D. & Reeve, H.K. *Am. Nat.* **131**, 602–610 (1988).
10. Page, R.E. *A. Rev. Ent.* **31**, 297–320 (1986).
11. Michener, C.D. & Brothers, D.J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 671–674 (1974).

Michael D. Breed is in the Department of Environmental, Population and Organismic Biology, University of Colorado, Boulder, Colorado 80309-0334, USA.

100 years ago

From *Nature* **38**, 39; 10 May 1888.

We have been enabled, through the kindness of Mr. Baker, to reproduce one of the photographs of the Forth Bridge, showing the successful completion of the lower portion of the first bay. The junction we have illustrated is the connection of the web of a lattice girder with one of its booms, but here the junction alone weighs as much as an ordinary iron railway bridge of 100 feet span.

Palaeobotany

Patterns before process

Chris Humphries

AFTER more than half a century of virtual stagnation in elucidating the phylogenetic relationships of the main groups of flowering plants, it became clear at a recent meeting* that a considerable change is now taking place. Exciting new results were reported on the study of major phylogenetic patterns and classification

includes plants as diverse as oaks, beeches, nettles and mulberries, has an extensive fossil record (see Figs 1 and 3) which stretches back to the Early Cretaceous and a major diversification in the Late Cretaceous (E.M. Friis, Swedish Museum of Natural History). Analysis of Late Cretaceous fossils shows that at least

platanoid fossils in the mid-Cretaceous floras suggests that these plants were important in the early divergence of the Hamamelidae and the Rosidae³ (see Fig. 3). The Juglandaceae (D.E. Stone, Duke University) the Ulmaceae (S. Manchester, Indiana University) and the Betulaceae (P.R. Crane, Field Museum, Chicago) all have rich fossil records dating back to the Upper Cretaceous. All three taxa exhibit a wealth of wholly fossil taxa, and it appears that modern genera made their first real appearances in the Tertiary⁴.

The character spectrum of the Hamamelidae seems to be very heterogeneous, giving the impression that the group is a series of broken-up fragments — the main orders being relictual survivors of an early phase of angiosperm evolution and remnants of a broad transitional group between the Magnoliidae and the Rosidae–Dilleniidae (F. Ehrendorfer, University of Vienna). But recent studies suggest that angiosperms are best rooted in the Magnoliales on the basis of comparisons with Bennettitales, Gnetales and Mesozoic seed ferns^{5–7}. The implication, then, is that the monosulcate Monocotyledonous groups are nested within a clade including the Nymphaeales, Piperales and possibly Aristolochiaceae and Lactoridaceae (see Fig. 2).

The primitively triaperturate groups appear to form a clade related either to the Winteraceae or to the primitive Laurales, and, if the tricolpate groups are monophyletic, then the Illiciales are basal and the Trochodendrales and the 'Hamamelidales' are related to the Ranunculidae rather than the Magnoliidae (M.J. Donoghue, University of Arizona and J.A. Doyle, University of California, Davis). It is therefore possible that

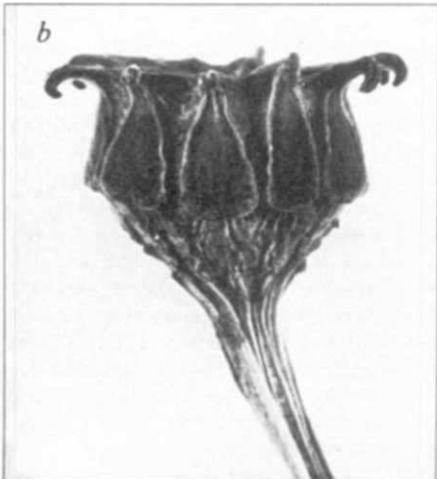


Fig 1 a, Fossil fruit of *Nordenskiöldia* from the Palaeocene of Montana, United States, compared with **b**, a young fruit of its extant relative *Trochodendron aralioides*. *Trochodendron* and its allies first appear in the fossil record during the mid-Cretaceous and are among the more primitive extant Hamamelidae. (Photographs courtesy of P.R. Crane).

within the families of the huge subclass called Hamamelidae and between the other subclasses of the angiosperms such as the Rosidae, Magnoliidae and Dilleniidae, which illustrate the remarkable progress made in palaeobotany over the past decade.

Because of the interaction between palaeontologists and neontologists, the main conclusion emerging from the meeting was that traditional subclass groupings (and indeed many other taxa) are paraphyletic grade groups destined to disappear from future phylogenetic reconstructions and taxonomic classifications. This detailed empirical reassessment of the Hamamelidae will make obsolete most current ideas about the pattern of angiosperm evolution. Thus, despite the argument recently aired in News and Views¹, that the processes by which angiosperms came to dominate lowland sediment-accumulating habitats are the most important for understanding angiosperm radiations, the study of systematic patterns, as emphasized by Lidgard and Crane², provides the critical results for elucidation of angiosperm diversification.

The Hamamelidae, which traditionally

two groups can be recognized: hamamelidalean forms, represented by tricolpate, reticulate pollen with several taxa assigned to the Platanaceae and one to the Hamamelidaceae; and juglandalean/myricalean forms represented by Normapolles-type pollen. The abundance of

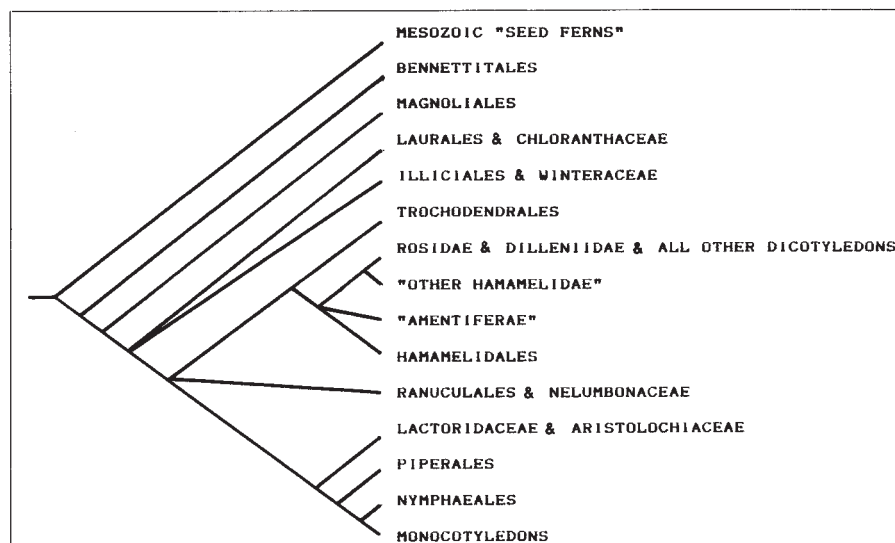


Fig. 2 A composite cladogram to show recent views as to the position of the groups of the Hamamelidae in the angiosperms and the seed plants. The Hamamelidae consist of at least three groups within a clade comprised of the Rosidae, the Dilleniidae and the majority of the higher dicotyledonous taxa.

* *Evolution, Systematics and Fossil History of Hamamelidae* Systematics Association, University of Reading, 22–25 March 1988.



Fig. 3 Reproductive axis of extinct *Platanus*-like plant from the Lower Cretaceous of Colorado, United States. *Platanus*-like plants were an important group in the mid-Cretaceous radiation of 'higher' dicotyledons and may have been close to the divergence of certain Hamamelidae and Rosiidae. (Photograph courtesy of P.R. Crane).

'Rosidae' are derived from a portion of the lower Hamamelids. Shared resemblances between the basal taxa of these two groups in various leaf, wood and floral characters (W.C. Dickson, University of North Carolina) shows the Hamamelidae to be paraphyletic (and hence not a group) without the inclusion of the Rosidae. Consequently, the traditional idea⁸ of considering them as having been independently derived from the Magnoliidae is false.

The origins and phylogenetic history of the Fagaceae are controversial and point out the difficulties inherent in attempts to differentiate the 'higher' Hamamelidae from 'primitive' Rosidae. Palynological characters and certain details of the flowers show that the Fagaceae could be more closely related to such families as Cunoniaceae and Brunelliaceae. The exact placement of groups such as the Fagaceae will eventually be helpful in resolving the precise Hamamelidae-Rosidae transition (K. Nixon, Cornell University).

It became clear during the meeting that certain taxa are considered wholly misplaced in the Hamamelidae. Most notable was the suggestion by H.D. Behnke (University of Heidelberg), on the basis of classification of sieve-tube plastids, that the Urticales do not have close relationships with other Hamamelidae but should be placed elsewhere. This view was expressed more precisely by C.C. Berg (University of Bergen) who thinks that the Urticales have sister-group relations with the Malvales and, more remotely, the Violales. The relationships have been obscured in the past by differing interpretations of reduced floral structures and many of the vegetative structures.

Extant hamamelids have a phenomenal array of breeding systems ranging from cross-pollination by specialized fig wasps (for example, W. Ramirez, University of Costa Rica, San Jose) to wind pollination in nettles (Friis). Although wind pollin-

ation is often considered as a recent, derived phenomenon, the presence of reduced floral structures in fossil hamamelids provides increasing evidence for the early evolution of anemophily in the Cretaceous. Hence, once again, refined palaeobotanical and neontological morphological studies of the Hamamelidae provide evidence for the reinterpretation of floral biology in early evolution. The work discussed will have considerable bearing on future studies on angiosperms. □

1. Cleal, C.J. *Nature* **331**, 304–305 (1988).
2. Lidgard, S. & Crane, P.R. *Nature* **331**, 344–346 (1988).
3. Friis, E.M., Crane, P.R. & Pedersen, K.R. *Nature* **320**, 163–164 (1986).
4. Friis, E.M., Chaloner, W.G. & Crane, P.R. (eds) *The Origins of Angiosperms and their Biological Consequences* (Cambridge University Press, 1987).
5. Crane, P.R. *Cladistics* **1**, 329–348 (1985).
6. Crane, P.R. *Ann. Mo. bot. Gdn.* **72**, 716–793 (1985).
7. Doyle, J.A. & Donoghue, M.J. in *Systematic and Taxonomic Approaches in Palaeobotany* (eds Spicer, R.A. & Thomas, B.A.) 177–198 (1986).
8. Cronquist, A. *An Integrated System of Classification of Flowering Plants* (Columbia University Press, 1981).

Chris Humphries is in the Botany Department, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK.

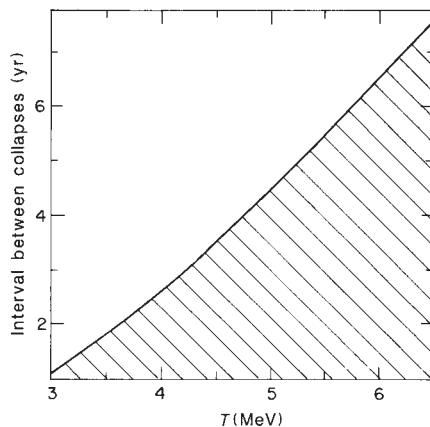
Neutrino astronomy

Supernovae leave their mark

John N. Bahcall

ASTRONOMERS have long believed that the direct radiation from a supernova is transient and stochastic, exemplified by a short intense burst of light that increases the optical brightness of a small region by a factor of a million or more and then fades quickly into obscurity. But Haxton and Johnson show on page 325 of this issue¹ that the neutrino bursts emitted in the collapse of supernovae in our Galaxy can contribute significantly to the average neutrino flux recorded in deeply buried mineral ores.

Their most interesting example involves



Limits that might be imposed by the molybdenum experiment according to Haxton and Johnson¹. The shaded area of assumed neutrino temperatures (T) from stellar collapse versus the collapse rate in years would be ruled out if all stellar collapses had the properties assumed and the neutrino absorption cross-sections were estimated accurately.

the recovery of technetium from the deep Henderson molybdenum mine in Colorado. The contribution of galactic supernovae to the accumulated abundance of ⁹⁷Tc produced over the past several million years could be as large as 40 per cent of the steady rate of production of the same isotope by solar neutrinos.

Neutrino astronomy, rewarded by the detection of neutrinos from the supernova 1987A, has been developing over the past

three decades, as physicists and chemists began to detect neutrinos emitted by the Sun. Neutrinos, essentially massless particles produced naturally by radioactive nuclei, interact only weakly with matter. These elusive particles can be detected by the reverse nuclear process — the transmutation of elements by neutrino absorption — or by rare collisions with electrons. Massive detectors, situated deep underground to screen out cosmic rays, are run for long periods to accumulate useful signals. Mined ores serve usefully as detectors that have been in place for millions of years.

The molybdenum-technetium connection discussed by Haxton and Johnson¹ is especially interesting because a group from Los Alamos National Laboratory is currently extracting ⁹⁸Tc and ⁹⁷Tc from the Henderson mine to perform a solar-neutrino experiment^{2,3}. The technetium that is produced has a half life that is measured in millions of years, so that the geochemical experiment effectively performs a time average over several million years of the neutrino flux from the Sun or galactic supernovae.

The isotope ⁹⁷Tc can be produced by neutrino absorption by ⁹⁷Mo and (indirectly) by ⁹⁸Mo, the primary isotope for solar-neutrino detection. Haxton and Johnson¹ estimate a combined production rate of ⁹⁷Tc from solar neutrinos of 3.9 solar neutrino units (SNU; 1 SNU = 10⁻³⁶ captures per target atom per second). They calculate a capture rate from supernova neutrinos of 1.6 SNU, which is produced predominantly via the indirect channel of ⁹⁸Mo. Thus supernovae may produce as much as 29 per cent of the ⁹⁷Tc that will be found in the Henderson molybdenum ore (40 per cent of the solar-neutrino rate).

The figure shows how limits can be imposed on the interval between galactic stellar collapses by the molybdenum-technetium experiment. The number of neutrinos absorbed depends upon the

assumed temperature (mean energy) of the neutrinos that are emitted from stellar collapses; values between 3 and 6 MeV cover nearly all theoretical stellar-collapse models. The shaded area could be excluded as a by-product of the Los Alamos solar-neutrino experiment.

Unfortunately, there is an important limitation in interpreting accurately both the supernova- and the solar-neutrino experiments. The neutrino absorptions of interest are all so-called Gamow–Teller transitions to relatively high excited states in the daughter nuclei. This particular kind of transition follows certain favoured selection rules named after the nuclear theorists George Gamow and Edward Teller. The only methods in use for estimating the rates of such transitions depend upon either theoretical models that cannot be derived from first principles or on laboratory measurements of the related proton-capture–neutron-emission (p,n) reactions.

For ^{98}Mo , the *a priori* theoretical calculation² of the solar-neutrino capture rate differs by a factor of 3 from the rate estimated with the aid of the (p,n) measurements⁴. Since the pioneering work of Goodman and colleagues on lighter nuclei^{5,6}, nuclear physicists have tended to prefer (p,n) measurements because these experiments nearly always give reasonable values in the moderate number of

cases in which the calibration of the technique was possible⁶. Neither method can provide a reliable estimate of the two- or three-standard-deviation uncertainty that is needed to interpret the observed astronomical rates. With the current lack of fundamental understanding, it could take about 100–1,000 measurements of different transitions in the same nuclear mass range with known (from β -decay) rates to establish a firm error estimate.

Nuclear physicists are faced with an important challenge. How can one calibrate accurately the systematic uncertainties, as well as the best estimates, for Gamow–Teller transitions in heavy nuclei? The results are needed to interpret the molybdenum–technetium experiments discussed by Haxton and Johnson, as well as some other important solar-neutrino experiments⁷. □

1. Haxton, W.C. & Johnson, C.W. *Nature* **333**, 325–329 (1988).
2. Cowan, G.A. & Haxton, W.C. *Science* **216**, 51–54 (1982).
3. Wolfsberg, K. et al. in *Solar Neutrinos and Neutrino Astronomy* (eds Cherry, M.L., Fowler, W.A. & Lande, K.) 196–202 (AIP Conf. Proc. No. 126, New York, 1985).
4. Rappaport, J. et al. *Phys. Rev. Lett.* **54**, 2325–2326 (1985).
5. Goodman, C.D. in *Solar Neutrinos and Neutrino Astronomy* (eds Cherry, M.L., Fowler, W.A. & Lande, K.) 109–120 (AIP Conf. Proc. No. 126, New York, 1985).
6. Taddeucci, N.T. et al. *Nucl. Phys.* **A469**, 125 (1987).
7. Bahcall, J.N. & Ulrich, R.K. *Rev. mod. Phys.* **60**, 297–372 (1988).

John N. Bahcall is at the Institute for Advanced Study, Princeton, New Jersey 08540, USA.

Visual texture perception

Features and spatial filters

Bela Julesz and Ben Kröse

A HUMAN observer has an uncanny ability to discriminate between two surfaces which are identical in colour or average luminance, but which differ in small-scale luminance variations — what is generally called 'texture'. We believe that texture discrimination is preattentive if the discrimination is immediate, and the observer does not need to scrutinize the textures to find differences in form (illustrated in Fig. 1). Preattentive texture discrimination is much simpler than form

perception, but adequately complex to challenge researchers in psychology, neurophysiology and machine vision. The central question is how texture elements can be computed from images of natural scenes, and how differences between these elements yield texture discrimination. There are two papers in this issue that address this question. Bergen and Adelson¹ on page 363 describe a very simple, size-tuned mechanism whose responses mimic the strength of human

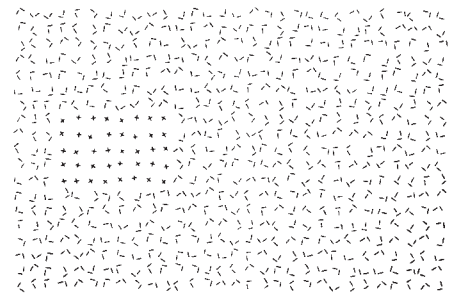


Fig. 1 In preattentive texture discrimination, aggregates of Xs embedded in aggregates of Ls segregate without scrutiny while the aggregates of Ts in Ls require element-by-element scrutiny. Note that the Xs, Ts and Ls are composed of two perpendicular line segments of equal brightness, width and length (from ref. 5). Bergen and Adelson¹ varied the relative sizes of Xs and Ls and find that they are easy to discriminate when they produce different size-tuned responses.

discrimination as the relative size of texture elements in adjacent textures is manipulated. And Voorhees and Poggio² on page 364 use a similar preprocessing stage as Bergen and Adelson, but their second stage is much more complex, performing statistical analysis over many attributes. Voorhees and Poggio's algorithm works not only for natural images, but also for some psychophysical findings — in particular, those of Bergen and Adelson.

How do the approaches described in this issue relate to current work on texture segregation? One way to characterize texture is to use a 'global' description. Within an adequately large window, the distribution of the raw intensity values is computed and a global (statistical) measure is used to describe the texture. We have shown^{3,4} that with such a global procedure many local properties are averaged out, so it is better to use local properties for the description of texture. Differences between such local primitives that are (for example, elongated blobs of given contrast, width, length or orientation) induce textural segregation, and these primitives are called textons^{3,5,6}. As stated by Bergen and Adelson¹, one major difficulty with this approach is that it is based on a verbal description of these primitives and

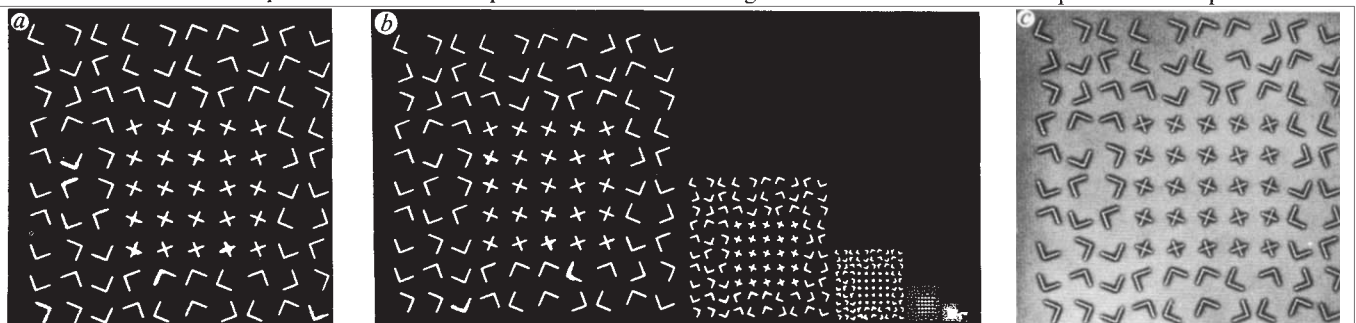


Fig. 2 *a*, Original texture pair; *b*, its decomposition by the laplacian pyramid, followed by squaring the output of the laplacian filters. Because of the limitations of photographic reproduction, only levels 0, 1, 2, 3, 4 and 5 are portrayed. The largest blob-contrast difference occurs at level 3. *c*, Reconstructed image, using the laplacian pyramid in reverse, but levels 2, 3 and 4 are changed to uniform grey. As can be seen, the resulting image is a highly discriminable texture pair, despite the fact that it lacks those levels (frequency bands) that contain the large energy differences in *b*.

that it would be much better to construct operators that extract the primitives invoked, "a task that has yet to be accomplished".

The chief importance of the new work of Bergen and Adelson is not so much that they have found a filter that can discriminate between line textures with and without crossings based on size differences, but rather their proposal that before searching for complex filters⁷ one should try simple linear filters followed by rectification for texture segmentation. Voorhees and Poggio in this issue describe an algorithm for the discrimination of textures based on attributes of blobs (for example, contrast, length, width or orientation). This algorithm explicitly specifies how the blobs and their attributes are extracted, while the blob-textons discovered by one of us^{3,6} were described only verbally.

Voorhees and Poggio compute their blobs by a centre-surround operator similar to that of Bergen and Adelson. As Voorhees and Poggio mention, an algorithm based on blobs is sensitive to the scale on which the blobs are determined. Some textures may contain many small blobs (high spatial frequencies), whereas others are built from a few large blobs (low spatial frequencies). Both authors in this issue^{1,2} use 'laplacian of gaussian' filters to extract blobs. To find an optimal blob size, they used filters of different standard deviation. One ingenious way to achieve this is to use a laplacian pyramid, where instead of filter size (standard deviation), the spacing between samples is increased in steps of two⁸.

We have used this technique for the decomposition of the original texture (Fig. 2a) into a series of band-passed filtered versions. Figure 2b shows the rectified output of the filters. The results show that for one particular filter size (level 3 in Fig. 2b), there is a large difference in 'blob contrast' between the textures (as found by Bergen and Adelson). This filter size is optimally suited to serve as an input for Voorhees and Poggio's statistical algorithm. As mentioned earlier, the blobs on this level can be considered as a band-pass filtered version of the original.

We wondered what the image would look like after we had removed the spatial frequencies in this band. Figure 2c shows the reconstructed image, leaving out the frequency band which caused the big energy differences, plus the bands adjacent to this frequency. As shown, segregation can be obtained. So, here is an

example of how to construct texture pairs that are easily discriminated, but respond equally to linear centre-surround filters at several scales. Despite our example, for which non-linear or more complex linear filters have to be used, the filters used in both reports in this issue are of great interest and should be tested on a wide variety of textures in order to see

how their responses parallel human discrimination. □

Bela Julesz is at AT&T Bell Laboratories, Murray Hill, New Jersey 07974, and is visiting professor in the Division of Biology, 216-76, California Institute of Technology, Pasadena. Ben Kröse is a research fellow in the Division of Biology, 216-76, California Institute of Technology, Pasadena, California 91125, USA.

Ocean Drilling Program

Early glaciation of Antarctica

Leg 119 shipboard scientific party*

THE timing and nature of the initiation of the Antarctic Ice Sheet is the subject of considerable discussion. Before Leg 119, the earliest known unequivocal Cenozoic glacial sediments were discovered in a Lower Oligocene sequence from the Ross Sea¹. Quartz grains of Eocene age from the subantarctic Pacific Ocean were inferred from their grain texture to be ice-rafted². Previous results obtained by Leg 113 in the Weddell Sea³ indicate that glaciation at sea level first occurred during the late early Oligocene on East Antarctica and during the late Miocene on western Antarctica. Our new results show that glaciation is present during the earliest Oligocene and possibly since the late Miocene.

The middle Cenozoic glacial record (starting 36–38 million years (Myr) ago) in the Ross Sea is the most complete record recovered from Antarctica¹. It records an early Oligocene phase of limited glaciation down to the coast and a late Oligocene phase of ice growth several times more extensive than that of the present. The earlier phase was tentatively attributed to local glaciation of the Transantarctic Mountains, the latter to full-scale development of an East Antarctic ice sheet.

We drilled at six sites on the Kerguelen Plateau in the southern Indian Ocean and five in Prydz Bay, East Antarctica (Fig. 1) to study the evolution of glaciation, to document preglacial environments in East Antarctica, and to study the late Cretaceous (about 65–90 Myr ago) and Cenozoic palaeo-oceanographic history of the

southern Indian Ocean. Prydz Bay lies at the mouth of the Lambert glacier–Amery ice shelf system which drains about a fifth of East Antarctica. This region was considered favourable to evaluate the glacial record and the first development of a full-scale ice sheet in East Antarctica.

Our results indicate that ice has been grounded on the continental shelf in Prydz Bay intermittently since earliest Oligocene (35 Myr) and possibly since late middle Eocene (42.5 Myr) times. Various

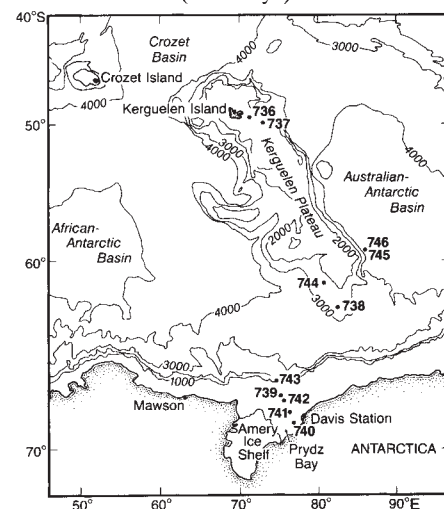


Fig. 1 Location of Leg 119 drill sites. Bathymetric contours shown in metres.

glacial sediment types, including lodgement tills and glaciomarine sediments, are present in sequences that are over 480 m thick. Most of the sediments are mixtures of clay, silt, sand and gravel — diamictites, characteristic signs of glaciation. Repeated loading of these sediments, possibly by ice, is indicated by their partially over-consolidated nature. Diamictites rich in metamorphic gravel and pebbles, derived from basement source areas, are present in thick prograding sequences (building out from the shoreline) extending for over 30 km on the Prydz Bay shelf.

Quaternary, Pliocene, upper Miocene and lowermost Oligocene diamictites at site 739 were dated by diagnostic diatom assemblages. Diatoms are sporadically

1. Bergen, J.R. & Adelson, E.H. *Nature* **333**, 363–364 (1988).
2. Voorhees, H. & Poggio, T. *Nature* **333**, 364–367 (1988).
3. Julesz, B. *Nature* **290**, 91–97 (1981).
4. Kröse, B.J.A. *Biol. Cybern.* **55**, 289–298 (1987).
5. Julesz, B. & Bergen, J.R. *Bell Sys. Tech. J.* **62**, 1619–1645 (1983).
6. Julesz, B. *Biol. Cybern.* **54**, 245–251 (1986).
7. Caelli, T. *Spat. Vis.* **1**, 19–30 (1985).
8. Burt, P.J. & Adelson, E.H. *IEEE Trans. Commun. COM-31*, 532–540 (1983).

*Co-Chief Scientists: John Barron (United States Geological Survey; USGS) and Birger Larsen (Tech. Univ. Denmark). ODP Staff Scientist: Jack G. Baldauf (ODP, Texas A&M Univ.). Also: Chantal Alibert (C.R.P.G. Dandoeuvre Cedex), Steven P. Berkowitz (Texas A&M Univ.), Jean-Pierre Caulet (Museum National d'Hist. Nat.), Steven R. Chambers (Stanford Univ.), Alan K. Cooper (USGS), Ray Cranston (Bedford Inst.), Wolfgang U. Dorn (Univ. Hawaii), Werner U. Ehrmann (Inst. Polar and Marine Res.), Richard Fox (Texas A&M Univ.), Greta Fryxell (Texas A&M Univ.), Michael J. Hambrey (Univ. Cambridge), Brian T. Huber (Ohio State Univ.), Christopher J. Jenkins (Univ. Sydney), Sung-Ho Kang (Texas A&M Univ.), Barbara Keating (Univ. Hawaii), Klaus W. Mehl (Ruhr-Univ. Bochum), Il Noh (Texas A&M Univ.), Gilles Ollier (Lamont-Doherty Geol. Observ.), Alan Pittenger (Texas A&M Univ.), Hideo Sakai (Toyama Univ.), Claudia J. Schroder (Dalhousie Univ.), Anders Solheim (Norwegian Polar Res. Inst.), Dean Stockwell (Univ. Rhode Island), Hans Thierstein (ETH), Bruce Tocher (Plymouth Poly.), Brian Turner (Univ. Cape Town) and Wuchang Wei (Florida State Univ.).

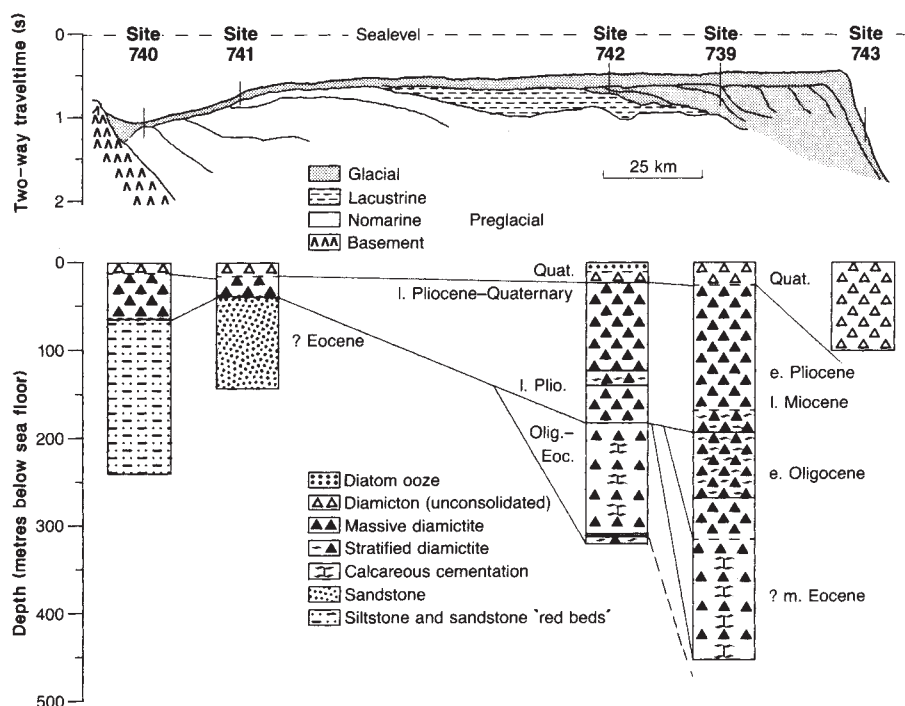


Fig. 2 Representation of the sediment facies recovered from Prydz Bay during Leg 119. Based on seismic and stratigraphic data collected during Leg 119. l, late; m, middle; e, early.

abundant and well preserved in the lowermost Oligocene unit and were not regarded as reworked by the shipboard diatom biostratigraphers. By correlating these with similar assemblages from the southern Kerguelen Plateau (site 744) dated magnetostratigraphically, we find that they date from at least 35.3 Myr ago (upper part of magnetic polarity chron 13). Unfortunately, magnetostratigraphy at site 739 is nondiagnostic. These earliest Oligocene diatom assemblages are underlain by 168 m of undated diamictite (Fig. 2), which is the lowermost sediment cored at site 739. Seismic stratigraphy reveals that these strata are younger than the basal 144 m of the section cored at site 742 (29 km to the south-east). The composite section of 300 m of diamictite, which underlies the dated lowermost Oligocene diamictite, contains sparse, possibly reworked but well-preserved calcareous nanofossil assemblages, which constrain their age to be middle Eocene to Oligocene.

Shipboard magnetostratigraphy indicates that the lower 144 m of this diamictite sequence at site 742 is dominantly of normal magnetic polarity but is interrupted by six short reversed events. The published magnetic polarity timescale below anomaly 13 shows only one interval in the lowermost Oligocene and Eocene characterized by such dominantly normal polarity separated by numerous brief reversals, namely the interval of anomalies 15 to 18 (37.24–42.73 Myr or late Eocene to latest middle Eocene). Within this interval, anomalies 17 and 18 (38.8–42.7 Myr) most closely resemble the magnetic polarity record at site 742.

Thus our drilling results suggest that a

large glacier complex reached the region of site 739 in Prydz Bay during earliest Oligocene time and possibly during late middle Eocene time (about 42.5 Myr), 6 Myr earlier than has been proposed for other Antarctic areas¹. The sediments also indicate that the outer limit of the ice front was beyond that of the present by at least 140 km and that full-scale ice-sheet development took place over East Antarctica by early Oligocene time or possibly earlier. This suggests that the early Oligocene glaciation recorded in the Ross Sea in West Antarctica was more than a local event and extended beyond the Transantarctic Mountains.

The series of late-Oligocene ice advances in the Ross Sea can be matched in Prydz Bay if we assume that ice advanced beyond site 739, removing sediment and causing the late Oligocene to middle Miocene hiatuses. Because of the sparse occurrence of datable microfossils and the incomplete sedimentary record, we cannot determine the full glacial chronology for Prydz Bay. But it is clear that the glacial depositional record extends through the early Oligocene, the late Miocene to early Pliocene and the late Pliocene to Quaternary times. Evidence of a Quaternary ice cover to the shelf edge was observed at site 743, which implies at least one advance of 170 km of the ice front. The entire Oligocene to Quaternary sediment record at site 739 gives evidence of the proximity of ice. Most of the time this ice was probably grounded. □

1. Barrett, P.J., Hambrey, M.J. & Robinson, P.R. in *Antarctic Earth Science Symposium Volume* (ed. Thompson, M.R.A.) (Cambridge University Press, in the press).
2. Margolis, S.V. & Kennett, J.P. *Am. J. Sci.* **271**, 1–36 (1971).
3. Leg 113 shipboard scientific party *Nature* **328**, 115 (1987).

Daedalus

Reagan's crystal

MOST chemical reactions release their excess energy as heat: random, disordered molecular motion. But, says Daedalus, consider an explosion initiated uniformly over one face of a perfect single crystal of a high explosive. Each molecule on that face will decompose identically, transmitting a pattern of molecular recoil to the next molecule below it, which will decompose in its turn.

Now a collision transmitted down a chain of spheres is 'self-focusing'. Even a highly oblique initial impact, as it travels down the chain, is rapidly aligned into an in-line impulse transmitted unchanged from sphere to sphere. Small 'thermal' disorders of velocity or position are also focused away. So a planar explosion wave in a single crystal should not scramble into thermal chaos, but should converge to a stable pattern handed on from layer to layer. Each type of decomposition molecule will spring off the reaction-face at a characteristic angle determined by the mechanism of reaction, so the exploding crystal will emit all its chemical energy as a fan of molecular beams. Each beam will consist of molecules of one type, travelling on accurately parallel paths at a velocity of many kilometres per second.

So DREADCO's intrepid chemists are cutting and polishing single crystals of selected high explosives, suspending them in vacuum, and launching planar shock waves into them by means of a laser pulse on one face. Once the velocity and direction of the resulting molecular beams has been determined (and apparatus strong enough to withstand their colossal impact on its walls has been constructed), Daedalus will look for technical applications. The most promising application is to rescue President Reagan's Strategic Defense Initiative from its current credibility crisis.

A parallel, high-velocity molecular beam, capable of being fired at a ballistic missile through thousands of kilometres of space, is a splendid space weapon. A cheap and simple single-crystal charge of high explosive, concentrating much of its energy in a deadly beam, is obviously far preferable to the nightmarishly complex and inefficient lasers and particle guns now lined up for the job. So many such charges could be economically accumulated in orbit that the crudest aiming software, blazing away furiously in true Wild West tradition, should hit most of the missiles in a flight by sheer random shotgun overkill. Best of all, a molecular beam (unlike a laser beam) is useless for Earth-bound warfare. Fired in the atmosphere, the most perfect single-crystal beam gun would simply explode in the conventional way, killing the gunner.

David Jones

DNA fingerprinting and 'scientific' whaling

SIR—We would like to report that the collection of small (200–300 mg) skin biopsies from free-ranging whales and their subsequent analysis by molecular genetic techniques is an alternative to 'scientific' whaling for addressing several important questions, including stock identification and population census.

Skin biopsy samples can be collected at a range of 10 to 30 m or more by firing a dart-tipped arrow from a cross-bow or compound-bow, air gun or rifle. As we have described previously¹, our dart is a metal cylinder 7 mm in diameter and 20 mm long. The arrow is tethered to a fly-casting reel and fired from an adjustable compound bow. Sample collection at sea is not dependent on refrigeration as the biopsy can be stored for extended periods in a salt-based preservative at ambient temperature.

Each sample yields 0.5–1.0 mg DNA, which is ample for many experimental analyses. The figure shows a 'DNA fingerprint' pattern and mitochondrial DNA (mtDNA) restriction digest from killer whale (*Orcinus orca*) DNA extracted from a skin biopsy collected at sea by the method described above. Human polycore minisatellite probes donated by Dr Alec Jeffreys (University of Leicester)² and a dolphin mtDNA probe donated by Dr Sarka Southern were used. DNA fingerprints from similar-sized skin samples taken from seven additional cetacean species are also shown, indicating a wide applicability for the technique.

One critical database for the conserva-

tion of cetacean populations is the census data currently gathered by sighting surveys, photographic identification of individuals, and marking studies where 'discovery tags' are fired into the whale to be recovered if the animal is later killed. DNA fingerprinting biopsy samples offers a minimally intrusive method with greater potential for accurate population census than current methods, although logistical considerations may make sighting surveys more practical for some species. Most vertebrate species show sufficient DNA fingerprint variation to allow individuals to be identified with vanishingly small probabilities of misidentification^{2,3}.

A fundamental concern for the long-term survival of these species is the maintenance of genetic variation, which is dependent on maintaining population numbers at a viable level. The impact of reduced variation on fitness for normally outcrossing species has been discussed extensively elsewhere⁴. Biopsy sampling has the potential to identify the genetic structure of populations and the dispersal and reproductive behaviour that contributes to the observed patterns of genetic variation. This can be achieved through paternity matching⁵ and the examination of various genomic components useful for describing variation at the population level, such as mtDNA and the ribosomal DNA gene family¹.

In a news article⁶ written at the time of the International Whaling Commission (IWC) meeting in Bournemouth last spring, Kathy Johnston described the controversy that arose when three member

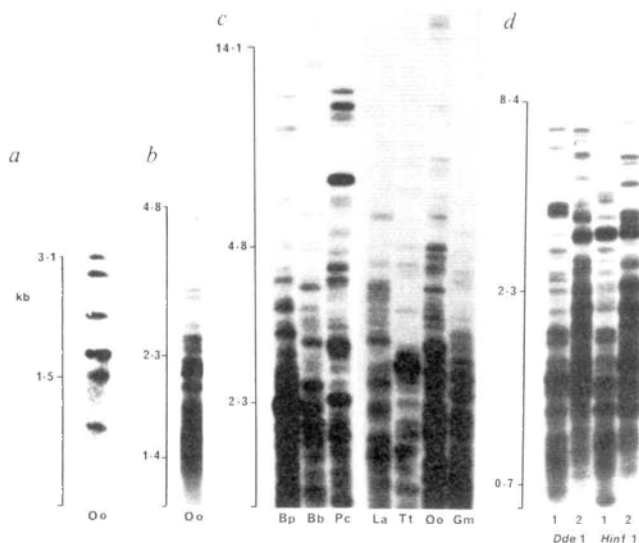
nations indicated their intention to take whales for scientific research in accordance with the IWC Convention, after agreeing to a three-year moratorium on whaling. There was disagreement within the IWC Scientific Committee over the feasibility of the scientific aims of Japan's intended catch, the size of which was later reduced to 300 minke whales before the fleet set sail in December. (A subsequent postal vote by the IWC delegates on a UK resolution that Japan refrain from allowing the catch to proceed registered 19 votes in favour, 6 against and 8 abstentions.)

For some time it has been a caveat of whale conservation and research that some aspects of cetacean biology can be investigated only using tissue samples from stranded animals or from a commercial or scientific catch. Indeed, the whaling industry has in the past provided the means to acquire important baseline data on age of sexual maturity, recruitment, population age structure, and various aspects of physiology. Given this database, however, our new priority should be to characterize accurately the size of populations and their genetic status. The technology needed to determine these fundamental population parameters without killing animals is at hand. This means that management policies and conservation measures may now be determined without the need to kill whales.

A. RUS HOELZEL
WILLIAM AMOS

Department of Genetics,
University of Cambridge,
Downing Street,
Cambridge CB2 3EH, UK

Mitochondrial DNA restriction pattern (a) and DNA fingerprint (b) from DNA extracted from a skin biopsy collected in the field, DNA fingerprints from five additional species (c) and variation in DNA fingerprint patterns with DNA from two individuals of the same species digested with two different enzymes (d). a, *O. orca* (Oo) mtDNA digested with *Hind* III and *Xba* I and probed with a *Cephalorhynchus commersonii* mtDNA probe labelled with ³²P. b, *O. orca* DNA digested with *Hinf* I and probed with a ³²P-labelled polycore human minisatellite clone (33.15)³. c, *Balaenoptera physalus*, *B. borealis*, *Physeter catodon*, *Lagenorhynchus acutus*, *Tursiops truncatus*, *O. orca*, and *Globicephala malaena* DNA digested with *Hinf* I and probed with the 33.15 clone. d, *Delphinus delphis* DNA from two individuals (1 and 2) digested with *Dde* I or *Hinf* I and probed with the 33.15 minisatellite clone.



1. Hoelzel, A.R. & Dover, G.A. *IWC Special Issue Series* (in the press).
2. Jeffreys, A.J., Wilson, V. & Thein, S.L. *Nature* **314**, 67–73 (1985).
3. Jeffreys, A.J. & Morton, D.B. *Anim. Genet.* **18**, 1–15 (1987).
4. Soule, M.E. (ed.) *Viable Populations for Conservation* (Cambridge University Press, 1987).
5. Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. *Nature* **317**, 818–819 (1985).
6. Johnston, K. *Nature* **327**, 548 (1987).

Supernova 1909A: a precedent for SN1987A?

SIR—The unusual light curve and the low peak luminosity of the type II supernova 1987A in the Large Magellanic Cloud (LMC) are attributed to the fact that the stellar progenitor, Sanduleak –69 202, was a blue supergiant, rather than a red one, when it exploded¹. The blue-supergiant nature of the progenitor is thought to have been connected with the relatively low abundance of heavy elements (low metallicity) in the LMC^{2,3}. Here we note the resemblance of the light curve of supernova 1909A to that of SN1987A. SN1909A appeared in the outskirts of the

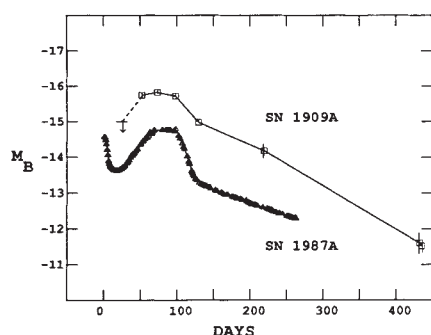


Fig. 1 Comparison of the SN1909A (ref. 4) and SN1987A (ref. 14) light curves. The time of the SN1909A explosion is assumed to be that which produces the best fit.

spiral galaxy Messier 101 (M101), where the metallicity is low. Observations are needed to determine whether massive stars are present at the site, and to search for radio emission analogous to that anticipated from SN1987A in future decades.

SN1909A (also called SS UMa) was discovered by Max Wolf at Heidelberg Observatory. Photometric coverage was fairly extensive, but because of the peculiar shape of its light curve and the lack of a spectral classification, SN1909A tends to be overlooked in supernova studies. In connection with their work on the extragalactic distance scale, however, Sandage and Tammann⁴ have discussed the SN1909A light curve, which is compared with that of SN1987A in Fig. 1. To put the light curves on an absolute scale, we have taken the distance modulus to SN1987A to be 18.5 magnitude (50 kpc), and the interstellar extinction in the blue band to be 0.8 mag. For SN1909A we have used a distance modulus⁵ of 28.8 mag (5.7 Mpc) and a blue extinction of 0.5 mag (a value characteristic of bright supergiants) in M101 (ref. 5).

The shape of the SN1909A light curve resembles that of SN1987A, although SN1909A appears to have been about a magnitude brighter. The brightness difference could conceivably just reflect the uncertainties in the adopted distance moduli (± 0.3 mag for SN 1987A, ± 0.5 mag for SN 1909A⁵ and extinctions, but a genuine difference in the amount of radioactive ^{56}Ni synthesized in the two explosions is perhaps more likely. Limits to the apparent magnitude of the SN1909A progenitor, $m_{\text{pg}} > 19.7$ mag in 1899 and 1908 (ref. 4), correspond to $M_{\text{pg}} > -9.6$, consistent with a supergiant progenitor (Sk -69 202 had $M_{\text{B}} = -7.0$).

In Fig. 2, the SN1909A light curve is compared to that of the 'linear' type II (II-L) supernova 1970G, also in M101, and to the composite 'plateau' type II (II-P) light curve. The comparison to SN1970G shows, independent of distance uncertainties, that SN1909A was less luminous at its peak than SN1970G. The shape of the SN1909A light curve is not at all like that of type II-L, nor like that of type I. Given

the limit of $M_{\text{B}} > -15.0$, about 45 days before it reached its peak, the SN1909A light curve also cannot be reconciled with type II-P.

SN1909A appeared 695 arcs s to the north-west of the nucleus of M101, well outside the optical image of the galaxy on an ordinary exposure, and just outside the image on a long exposure⁴. If the progenitor of SN1909A was a supergiant, then other massive stars, and perhaps nebular emission, would be expected near the supernova site. If it had been at the distance of M101, Sk -69 202 would have had $m_{\text{pg}} = 22.3$, well within the range of ground-based telescopes. Consistent with the possible parallel between SN1909A and 1987A, the metallicity at the site of SN1909A is low. In M101, a nearly face-on Sc galaxy decreases with increasing distance from the galactic centre; at the

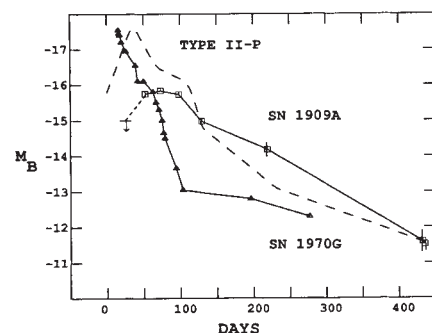


Fig. 2 Comparison of the light curve of SN1909A (ref. 4) to that of the type II-L SN1970G (ref. 15) in the same galaxy, and to the composite type II-P light curve¹⁶. The peak absolute magnitude of type II-P is taken to be the same as SN1970G.

outlying position of SN1909A the oxygen abundance is four times lower than in the Sun (Fig. 1 of ref. 6) — about the same as in the LMC.

During the first weeks after its explosion, SN1987A was a weak radio source, evidently owing to a hydrodynamical interaction between matter ejected by the supernova and a low-density circumstellar shell resulting from a moderate-velocity ($\sim 500 \text{ km s}^{-1}$) stellar wind during the blue-supergiant phase of Sk -69 202 (ref. 7). But Sk -69 202 is likely to have gone through a red-supergiant phase thousands or tens of thousands of years before it exploded⁸⁻¹⁰, at which time a slow ($\sim 10 \text{ km s}^{-1}$) wind may have created a dense shell. Stronger radio emission is anticipated when, perhaps after decades, the ejected matter reaches the dense shell¹¹. SN1909A, being decades more advanced than SN1987A, might provide a radio preview; unfortunately, SN1909A also is over 100 times more distant. Upper limits to the radio emission from SN1909A have been established at 2 mJy at 20 cm (in 1971)¹² and 4 mJy at 4 cm (in 1974)¹³, but a Very Large Array search could now go more than a factor of 10 deeper.

F. P. Israel *et al.*¹⁷ also established a 2-

mJy upper limit in 1971, and provided a position (epoch 1950) for SN 1909A.

TIMOTHY R. YOUNG

DAVID BRANCH

Department of Physics and Astronomy,
University of Oklahoma,
Norman,
Oklahoma 73019, USA

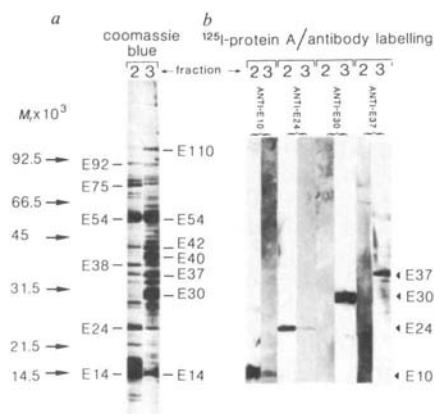
1. Woosley, S.E., Pinto, P.A. & Ensmann, L. *Astrophys. J.* **324**, 466–489 (1988).
2. Arnett, W.D. *Astrophys. J.* **319**, 136–142 (1987).
3. Hillebrandt, W., Hoflich, P., Truran, J.W. & Weiss, A. *Nature* **327**, 597–600 (1987).
4. Sandage, A. & Tammann, G.A. *Astrophys. J.* **194**, 223–243 (1974).
5. Humphreys, R.M. & Aaronson, M. *Astr. J.* **94**, 1156–1168 (1987).
6. Evans, I.N. *Astrophys. J.* **309**, 544–552 (1986).
7. Chevalier, R.A. & Fransson, C. *Nature* **328**, 44–45 (1987).
8. Woosley, S.E. *Astrophys. J.* (in the press).
9. Saio, H., Kato, M. & Nomoto, K. *Astrophys. J.* (in the press).
10. Kirshner, R.P. *ESO Workshop on SN1987A* (ed. Danziger, I.J.) 121–131 (European Southern Observatory, Munich, 1987).
11. Chevalier, R.A. *ESO Workshop on SN1987A* (ed. Danziger, I.J.) 481–494 (European Southern Observatory, Munich, 1987).
12. de Bruyn, A.G. *Astr. Astrophys.* **26**, 105–112 (1973).
13. Brown, R.L. & Marscher, A.P. *Astrophys. J.* **220**, 467–473 (1978).
14. Catchpole, R.M. *et al. Mon. Not. R. astr. Soc.* (in the press).
15. Barbon, R., Ciatti, F. & Rosino, L. *Astrophys. J.* **29**, 57–67 (1973).
16. Doggett, J.B. & Branch, D. *Astr. J.* **90**, 2303–2311.
17. Israel, F. P., Goss, W. M. & Allen, R. J. *Astr. Astrophys.* **40**, 421 (1975).

Import receptor in chloroplast envelope

SIR—In their recent article¹, Pain *et al.* identified an integral membrane protein of the chloroplast envelope as a receptor for protein import into the chloroplast stroma. The anti-idiotypic antibody approach they used to demonstrate that the import receptor is found in contact sites between the outer and the inner membrane of the chloroplast envelope is convincing. In contrast, the evidence concerning the identification of the import receptor raises several questions.

First, two chloroplast proteins, the large subunit of the ribulose biphosphate carboxylase (Rubisco) and a membrane-bound polypeptide of relative molecular mass 30,000 (30K) react with the anti-idiotypic antibody. Non-specific binding of the antibody with the large subunit of the Rubisco is not excluded by the authors, but they do not invoke such a possibility for the 30K polypeptide. It is well known²⁻⁴ that thylakoid and envelope membranes contain several 30K polypeptides. The most interesting in the envelope membrane is the phosphate translocator⁵ (see figure). It is surprising that the authors did not mention this major envelope protein or discuss the possibility that the protein immunoprecipitated in the chloroplast membrane fraction could be the phosphate translocator.

The procedure used by the authors to fractionate the chloroplast membranes by flotation allowed the separation of an envelope fraction containing both the



Electrophoretic and immunochemical analyses of envelope polypeptides in membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts (Fig. 3 from ref. 4). *a*, coomassie blue staining of envelope polypeptides separated by SDS-PAGE with a 7.5–15 per cent acrylamide gradient; *b*, autoradiography of antigen-antibody ^{125}I -protein A complexes after electrophoretic transfer to nitrocellulose sheets. Fractions 2 and 3 correspond to membrane fractions enriched in outer and inner envelope membranes, respectively. Antibodies to E10 and E24 are against outer envelope membrane proteins, whereas antibodies to E30 and E37 are to inner envelope membrane proteins; E30 is the phosphate translocator.

outer and the inner envelope membranes. For instance, this fraction was used by Cline *et al.*³ as a source of both outer and inner membranes for further separation of each membrane. Unfortunately, Pain *et al.* did not investigate whether their membrane fraction contained only the outer-envelope or both envelope membranes. The presence of bound pS in this fraction was expected but this does not exclude the possibility that some inner membrane (and therefore some phosphate translocator) is also present. To identify the receptor, the authors cite as evidence: (1) that a major 30K polypeptide exists in envelope membranes after silver-staining (their Fig. 6a); (2) that this protein is also the major iodinated polypeptide of the envelope fraction (their Fig. 6c); and (3) that this protein precipitated in presence of the antibody (their Fig. 6c). The only integral envelope membrane protein following the first two criteria is indeed the phosphate translocator, which is the major protein in the inner envelope membrane (see figure). There is as yet no proof that the import protein is a major envelope component. Because the antibody is able to immunoprecipitate major components of the fractions analysed (their Fig. 5), one can question whether the major protein immunoprecipitated in their Fig. 6c is indeed the import receptor. We believe that the most likely candidates are the minor bands indicated by asterisks in Fig. 6c or a minor protein which coprecipitates with the phosphate translocator.

We believe that more work is needed before it can be concluded that the import receptor of the chloroplast envelope is a major 30K integral membrane protein. Envelope membranes contain many different proteins involved in the biosynthesis of typical plastid components (such as galactolipids, phosphatidylglycerol, prenyl-quinones) and in the exchange of metabolites between the cytosol and the stroma. Only a few of these proteins have been identified. The main reason for this is that large amounts of envelope membranes have to be purified from chloroplasts, a difficult procedure.

JACQUES JOYARD

ROLAND DOUCE

*Laboratoire de Physiologie Cellulaire
Végétale UA au CNRS n° 576,
Département de Recherche Fondamentale,
Centre d'Etudes Nucléaires et Université
Joseph Fourier,
85 X, F-38041 Grenoble cedex,
France*

1. Pain, D. *et al.* *Nature* **331**, 232–237 (1988).
2. Joyard, J. *et al.* *J. biol. Chem.* **257**, 1095–1101 (1982).
3. Cline, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 3595 (1981).
4. Block, M.A. *et al.* *J. biol. Chem.* **258**, 13273–13280 (1983).
5. Flugge, U.I. & Heldt, H.W. *FEBS Lett.* **68**, 259 (1976).

PAIN AND BLOBEL REPLY —We agree with Joyard and Douce that a chloroplast fraction, when analysed by SDS-PAGE, yields a major 30K band which at high resolution consists of several polypeptides and that more work will be necessary to definitively establish which of these is the import receptor.

Since publication of our paper¹, we have prepared rabbit antibodies against this major 30K band excised from an SDS gel. These antibodies are able to agglutinate intact chloroplasts and, like our anti-idiotypic antibodies, yield distinct immunofluorescent staining in patches in the periphery of chloroplasts. Moreover, monovalent F_{ab} fragments of these antibodies inhibit import. These data, which are yet to be published, thus provide additional support for our previous conclusion that a 30K polypeptide of the chloroplast envelope that we found to be immunoreactive with the described anti-idiotypic antibodies is likely to be the import receptor (or part of it).

DEBKUMAR PAIN

GUNTER BLOBEL

*Rockefeller University,
1230 York Avenue,
New York, New York 10021-6399, USA*

How to avoid 'useless' radiocarbon dating

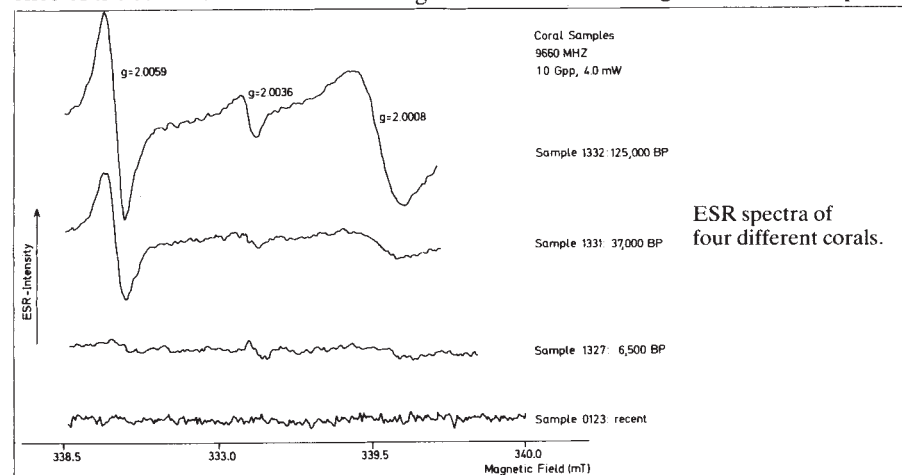
SIR—I wish to propose a quick and inexpensive method of screening samples to determine their suitability for radiocarbon dating. This would reduce the large amount of money that is wasted by dating samples that in fact lie either completely beyond the limit of ^{14}C dating, or within the 'critical range' of 20,000–50,000 years before present (BP) in which ^{14}C dating is possible but inaccurate.

For example, 51 of the 272 shell and coral ^{14}C datings published in 1983 in volume 25(3) of the journal *Radiocarbon* were 'useless'. Assuming an average cost of \$250 for each dating, this represents a waste of \$12,750 of research funds that could have been spent more fruitfully. Pre-screening would also help to avoid controversies such as that about the existence of the so-called Mid-Wisconsin high-

level (~35,000 yr BP) which is based solely on the misinterpretation of ^{14}C dates.

In many cases, the stratigraphical context of a particular sample does not make it clear whether or not the sample lies within the optimal ^{14}C dating range. I suggest that samples are screened for their approximate age range by means of electron spin resonance (ESR) which can date various paramagnetic materials such as quartz, calcite, aragonite, feldspar and zircon¹.

Unfortunately, the material analysed most frequently by ^{14}C dating — charcoal — is unsuitable for ESR dating, but shells and corals are common examples of suitable calcitic and aragonitic material. In aragonitic fossils there are three important ESR signals that increase with time^{2,3}. As shown in the figure, in contemporary



samples there are no such signals, in Holocene samples the peaks are very small, but in Pleistocene samples they are quite pronounced.

Although the procedure of ESR dating of samples is tedious and involves many steps, including determination of the ESR spectra, screening requires only this single step. Measuring an ESR spectrum takes only 5–10 min. If another 5–10 min is allowed for sample preparation time (cleaning and grinding the sample of at least 20 mg) hundreds of research dollars can be saved by 10–20 min work. Of course an ESR spectrometer is necessary, but nearly every university has one.

Badly preserved or contaminated samples are not suitable for the ESR screening method (and are probably not suitable for ^{14}C analysis either) but in general, for reasonably well preserved samples, it should be an extremely effective technique. Screening could significantly reduce the number of questionable ^{14}C dates of carbonates. If screening indicates an age that is beyond the limit of ^{14}C dating, then another dating method can be substituted.

ULRICH RADTKE

Geographisches Institut,
Universität Düsseldorf,
D-4000 Düsseldorf 1, FRG

1. Hennig, G.J. & Grün, R. *Quat. Sci. Rev.* **2**, 157–238 (1983).
2. Radtke, U. *et al. Nucl. Tracks* **10**, 879–884 (1985).
3. Radtke, U. *et al. Quat. Res.* (in the press).

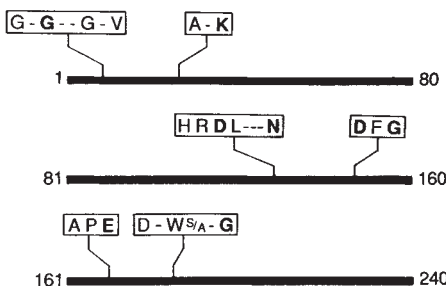
Identification of protein kinases by computer

SIR—I would like to add a cautionary note to the recent letter of Bairoch and Claverie¹ about recognition by computer of eukaryotic protein kinases by sequence patterns. These authors claim that two patterns can be used to identify protein kinases and discriminate them from other proteins.

Here, I make three points. First, both these patterns are absent from the protein kinases casein kinase II (ref. 2) and *nim* 1 (ref. 3). Second, as the authors kindly informed me when I requested its complete sequence, the TMS1 protein that they predicted to be a protein kinase on the basis of one of these patterns¹ is, in fact, tryptophan 2-monooxygenase⁴. Third, eukaryotic protein kinases can easily and accurately be identified by other means.

Such identification may, if necessary, involve initial sequence comparison with the protein databases (for example, using a program based on the algorithm of Wilbur and Lipman⁵), or direct comparison with protein kinases (for example, using a program based on the algorithm of Needleman and Wunsch⁶). Initial 'matches' of unknown or candidate proteins to known eukaryotic protein kinases must then be examined ('manually' or 'by

eye') for the presence of the motifs shown in the figure, which are derived from well-known similarities^{7,8}. The amino-acid residues shown in bold have, so far, been found to be totally conserved and, although the other residues can vary, in known protein kinases (including casein kinase II and *nim* 1) no more than a single deviation per motif has been observed. In practice, the presence of other semi-conserved residues, such as the hydrophobics in Bairoch and Claverie's patterns, in the regions of these motifs further assists unambiguous identification.



Highly conserved motifs in eukaryotic protein kinases. The average positions of these in the approximately 240-amino-acid protein kinase domain is indicated (see refs 7 and 8).

It may seem self-evident that all these motifs should be taken into account, but there has been a tendency in some quarters to focus on the G-G-G of the most amino-terminal motif (equivalent to Bairoch and Claverie's pattern 2) because of its likely location at the nucleotide-binding site of protein kinases (but also of certain other enzymes⁹). Although the functions of the other motifs (except A-K) are unknown, their importance can easily be imagined if one considers the locations of conserved aspartate and asparagine residues¹⁰ in the three-dimensional structure of the GTP-binding protein EF-Tu^{11,12}. The extreme conservation in diverse prokaryotic and eukaryotic phosphotransferases of the Asp residues represented in the protein kinase motifs HRDL---N and DFG has already been pointed out by Brenner¹³.

Computers are indispensable in modern biology. A molecular biological perspective is also helpful.

DAVID P. LEADER

Department of Biochemistry,
University of Glasgow,
Glasgow G12 8QQ, UK

1. Bairoch, A. & Claverie, J.-M. *Nature* **331**, 22 (1988).
2. Saxena, A. *et al. Molec. cell. Biol.* **7**, 3409–3417 (1987).
3. Russell, P. & Nurse, P. *Cell* **49**, 569–576 (1987).
4. Gielen, J. *et al. EMBO J.* **3**, 835–846 (1984).
5. Wilbur, W.J. & Lipman, D.J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).
6. Needleman, S.B. & Wunsch, C.D. *J. molec. Biol.* **48**, 443–453 (1970).
7. Takio, K. *et al. Biochemistry* **23**, 4207–4218 (1984).
8. Hunter, T. & Cooper, J.A. *Rev. Biochem.* **54**, 897 (1985).
9. Möller, W. & Amons, R. *FEBS Lett.* **186**, 1–7 (1985).
10. Dever, T.E., Glynn, M.J. & Merrick, W.C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1814–1818 (1987).
11. Jurnak, F. *Science* **230**, 32–36 (1985).
12. La Cour, T.F.M. *et al. EMBO J.* **4**, 2385–2388 (1985).
13. Brenner, S. *Nature* **329**, 21 (1987).

Lod score or log-likelihood?

SIR—The publication of another paper¹ in *Nature* which uses 'Lod scores' to describe the analysis of human linkage data prompts me to wonder whether it might not be an opportune moment for linkage workers to abandon their idiosyncratic statistical terminology so that others can better follow their analyses.

Their Lod score is simply the log-likelihood to the base 10, standardized as is customary by the addition of a constant, in this case such as to make the log-likelihood zero at a recombination fraction of one-half. The phrase is confusing because in standard likelihood terminology it is neither a lod nor a score: lod ('log-odds' was first defined by Barnard² in terms of natural logarithms, and score is now reserved for the first derivative of the log-likelihood^{3,4}. Indeed, the word odds itself is nowhere else used in this 'backward' sense introduced by Barnard, the common name being the likelihood ratio. The paper also uses the term 1-Lod confidence interval although it is not a confidence interval in Neyman's sense, as has often been pointed out⁵.

I imply no criticism of linkage workers by suggesting that a change to conventional terminology might be timely. On the contrary, linkage analysis in man has been a valuable proving-ground for the use of likelihood as a primary statistical tool, and the terminology introduced by Smith⁶ and Morton⁷ probably helped to make likelihood methods acceptable (even though that was not the intention of either author at the time⁶). But just as physicists eventually abandoned probable errors in favour of standard deviations, so human geneticists might with advantage abandon Lod scores in favour of the customary log-likelihoods to the base *e*. And rather than using the phrase 'confidence interval' in a non-Neyman sense, they might consider my phrase "support interval"⁴ which was specifically coined for likelihood use.

There is, of course, no difficulty in adjusting the standard computer programs for linkage so as to work with natural rather than common logarithms (some already offer this option), and by changing the terminology at the same time as changing the logarithmic base there will be no risk of confusion.

A.W.F. EDWARDS

Department of Community Medicine,
University of Cambridge,
Fenner's, Gresham Road,
Cambridge CB1 2ES, UK

1. Wasmuth, J.J. *et al. Nature* **332**, 734–736 (1988).
2. Barnard, G.A. *J. R. statist. Soc. B11*, 115–149 (1949).
3. Bailey, N.T.J. *Introduction to the Mathematical Theory of Genetic Linkage* (Clarendon, Oxford, 1961).
4. Edwards, A.W.F. *Likelihood* (Cambridge University Press, 1972; 1984).
5. Smith, C.A.B. *Ann. hum. Genet.* **50**, 293–311 (1986).
6. Smith, C.A.B. *J. r. statist. Soc. B15*, 153–192 (1953).
7. Morton, N.E. *A. J. hum. Genet.* **7**, 277–318 (1955).

Too many books spoil the cook

Frank Gannon, John Neilan and Richard Powell

Guide to Molecular Cloning Techniques. Edited by Shelby L. Berger and Alan R. Kimmel. Academic: 1987. Pp.813. Hbk \$89, £56; spiral bound \$45, £28.50.

Current Protocols in Molecular Biology. Edited by Frederick M. Ausubel et al. Greene Publishing/Wiley: 1987. Pp.676. Pbk and loose-leaf format \$140, £96.55 (quarterly update service \$95, £66.50 per year).

Basic Methods in Molecular Biology. By Leonard G. Davis, Mark D. Dibner and James F. Battey. Elsevier: 1986. Pp.388. Spiral bound Dfl 126, \$39.50.

Gene Cloning and Analysis: A Laboratory Guide. Edited by G.J. Boulnois. Blackwell Scientific: 1987. Pp.144. Pbk £14.50.

DNA Cloning: A Practical Approach, Volume III. Edited by D.M. Glover. IRL: 1987. Pp.254. Hbk £26, \$47; pbk £17, \$32.

PUBLISHERS obviously think that there is a big demand for laboratory manuals describing genetic engineering techniques, and one doesn't need market research to agree with them. The success of genetic engineering in deepening our knowledge of how biological systems are controlled, expressed or modified, and in providing products of pharmaceutical or diagnostic importance, has been well highlighted. This cornucopia has resulted from what is in essence a collage of techniques. Every laboratory working in the area needs a printed guide in which the various techniques are brought together and in which novel developments are clearly presented.

This need was first recognized in the late 1970s by biochemists who wanted tuition in how to use lambda, and by microbiologists who wished to use the enzyme reverse transcriptase. When practical courses were organized the time came to write down detailed protocols, and such courses are still the source of some manuals (including one of those reviewed here). In 1982 came the appearance of *Molecular Cloning* by Maniatis, Fritsch and Sambrook, a publication that for many people has become the practical bible in molecular biology.

But things change. The methods themselves evolve; suppliers of reagents provide increasingly sophisticated back-up services; and there are now many kits, which come complete with detailed protocols and allow even an inexperienced operator to carry out complex experiments. Discussion of the merits of kits against those of laboratory-derived methods is not appropriate here. But given the high cost of kits, those who think the prices of some (in particular one) of the publications under review are too steep might think again.

Two of the books, *Guide to Molecular Cloning Techniques* and *Current Protocols in Molecular Biology*, have ambitions to replace Maniatis et al. (a new edition of which is due in the next few months) as the laboratory manual. Berger and Kimmel's

multi-author *Guide* comes from Academic Press's excellent *Methods in Enzymology* series. It is comprehensive, with 10 sections covering 74 topics. Each author writes with clear practical knowledge of the method under discussion and appears to have been given the freedom of space and style to present the method in an appropriate way. Thus "The Cloning of cDNA into λ GT10 or λ GT11" is treated in a step by step manner, with detailed subsections, whereas "Gene Cloning Based on Long Oligonucleotide Probes" is a short chapter in which the aim is to highlight the usefulness of this approach without repeating the methodological details covered elsewhere.

In all chapters there is an excellent balance between discussion of the scientific basis of the method and description of the actual procedure. References to examples in the literature which illustrate the use of the method are plentiful. Each chapter is adequately cross-referenced, and there is both an index and a "process guide" in which specific methods are grouped together.

Guide to Molecular Cloning Techniques passed every test to which we subjected it. In addition to a comprehensive range of basic techniques, it provides detailed descriptions of difficult aspects of routine experiments (for example preparation of the buffer for terminal transferase) and of up-to-date methods such as the detection of *in vitro* protein-DNA interactions or the use of RNA transcripts in S_1 nuclease mapping. We recommend it highly, especially for a laboratory in which there is existing expertise in the area. Some of our colleagues who had no previous experience in molecular biology techniques would have preferred a more clear-cut presentation, as found for example in *Current Protocols in Molecular Biology*. But that, no doubt, would require a complete change of format, and the compact book as presented has its advantages.

Current Protocols in Molecular Biology is, to quote the publisher, "a manual with a difference". As well as in a paperback

edition, it comes in a loose-leaf binder with quarterly updates (at a price). There are obvious benefits in this second format, given the continual elaboration of existing techniques and the development of new ones. The core manual is already very large, requiring 70×29 cm of bench space with the 670 pages stacked 10 cm high. No doubt the publishers have studied the structural engineering aspects required to allow for a further 400 pages to be added annually.

The core manual contains 10 topic sections; these are readily found by the aid of titled dividers which protrude on the side of the page. Again this is a multi-author work. It has been well edited, and great care is given to presentation. Each topic has an introduction with key references. The method is then described, step-by-step, with useful comments on the step included in italics. Techniques which are frequently used are also presented in a shortened table form. Support protocols which cover associated aspects of the method or alternative procedures follow, and their presence close to the basic method is very convenient. Finally, each description of a method ends with a "Commentary" section which includes background information, an indication of the crucial steps in the protocol and a guide to trouble-shooting.

In all, this is a thoroughly satisfying compilation, very user-friendly and well laid-out. The core manual contains a wide range of methods, including some which were new to us (for example a method for plasmid preparation using a 96-well microtitre plate) and the promise of updates will ensure that the book will remain popular for some time.

It is perhaps rather unfair to include Davis, Dibner and Battey's *Basic Methods in Molecular Biology* in this round-up, because it appeared about two years ago. The book covers a wide range of methods in a brisk manner, but there is little background information provided or reference to examples from the literature in which the procedure has been used. In this way it is not as intellectually enriching as Berger and Kimmel. The procedures are again presented in a stepwise fashion, with footnotes to expand on some aspects, but overall this is a less clear and helpful manual than *Current Protocols in Molecular Biology*.

This book is also a reminder that it is difficult to keep laboratory manuals up to date — thus, for example, although it contains an extensive collection of protocols, it appears to have been written before oligolabelling of DNA probes became popular. We understand that a second edition is currently being prepared, which will appear in about a year's time.

The aims of *Gene Cloning and Analysis*, edited by Boulnois, are more modest. The

book is based on a course run at the University of Leicester, and the methods are selective but well described, as one would expect from its provenance. Little effort, however, has been made to make the contents of more general use. There is no index and some methods occur in unexpected places; for example the transformation of *E. coli* is found in the section on cDNA cloning. It is always useful to see how experiments are performed in other laboratories, but those buying *Gene Cloning and Analysis* should be aware that it contains a limited range of techniques.

The last book, Vol. III of Glover's *DNA Cloning*, is a recent addition to the successful *Practical Approach* series published by IRL Press (Vols I and II were reviewed in *Nature* 317, 679; 1985). Ten topics are covered in the new volume, each of them by an expert in the area. There are two chapters on the uses of cosmids and three on different aspects of eukaryotic polypeptides expressed in *E. coli*. The subjects of other contributions include the use of yeast and mammalian cells as hosts for the expression of foreign genes, retroviral vectors and transgenic mice, and there is also a timely and up-to-date chapter on the generation and use of RNA probes. The difference between this book and the others reviewed here is that it addresses very specific and occasionally esoteric topics, and does not claim to be a comprehensive laboratory manual. Nonetheless, the book and its predecessors make very useful companions in the laboratory.

It is clear that the manual needs of the molecular biologist are now as well catered for as those of DIY enthusiasts. Such manuals should never replace the drive inherent in a scientist to do things quicker and better; an unquestioning dependence on the protocols of others can stunt one's intellectual growth. For all that, some of these books are now welcome fixtures in our laboratory, and their appearance has stopped forever a project to publish our own manual of gene cloning techniques. □

Frank Gannon is Acting Director of Bio-Research Ireland's National Diagnostic Centre, Galway, and a Lecturer in the Department of Microbiology, University College, Galway, Ireland. John Neilan and Richard Powell are Postdoctoral Fellows in his laboratory.

• Molecular biology manuals continue to appear — recently published by Martinus Nijhoff is *Plant Molecular Biology* edited by Stanton B. Gelvin and Robert A. Schilperoort, a loose-leaf manual to which further sections can be added as they are published. Price is Dfl.110, \$52.50, £32.50. Also, the second edition of *A Practical Guide to Molecular Cloning* by B. Perbal is to be issued by Wiley in June, with updated descriptions of sequencing, *in vitro* expression of cloned genes, and use of computers for the study of nucleic acids. Price is hbk £90, \$99.50; pbk £42.50, \$49.50. For review see *Nature* 316, 222 (1985).

Material change

John R. Barker

GaAs Devices and Circuits. By Michael Shur. Plenum: 1987. Pp.670. \$90, £50.

THERE has long been a popular belief that gallium arsenide (GaAs) will eventually replace silicon as the basic material for advanced integrated-circuit technology. The argument is usually based on the assumption that industry needs ever-faster electronic devices and the fact that gallium arsenide is intrinsically a much faster material than silicon. By 'faster' it is meant that the electrons in gallium arsenide may be accelerated more rapidly and achieve higher limiting speeds than in silicon. The speed difference is set by fundamental material properties such as the energy-band structure and the details of electron-phonon scattering. There are signs now of rapid growth in commercial GaAs integrated-circuit technology, but the ultimate competition with silicon is not one just of speed.

Silicon has achieved its present supremacy by a scaleable technology which can accurately produce digital logic and memories which are not only fast but which have high-integration densities. The key is the existence of a well-matched insulator, silicon dioxide, which provides the basis for the highly controllable metal-oxide-semiconductor field-effect transistor (MOSFET) technology. Unfortunately there is no comparable insulator for gallium arsenide, and the fundamental device technology is therefore quite different. With gallium arsenide, the basic field-effect transistor action is achieved by using the potential barrier at a metal-semiconductor interface to control the size of the conducting region in the device channel; this is the basis of metal-semiconductor field-effect transistor (MESFET) technology. At first sight this alternative technology should not be a problem, and record-breaking oscillators were being produced by 1.0 µm gate GaAs MESFETs as long ago as 1970.

On the other hand, for MESFET-based digital logic or memories it is vital to have extremely accurate control of both the layer thicknesses and the doping distribution within the conducting layers. Compared to silicon, the subsequent materials technology is hampered by the lower solubility of impurities. The lower chemical stability of gallium arsenide puts a much greater restriction on the range of processing techniques and temperatures that can be used. The roots of the problem may thus be partly traced to the long learning curve in the underlying materials science of gallium arsenide and the restrictions that has placed on the device structures. However, the past two years have

seen commercial production of GaAs integrated circuits, including gate arrays, and the announcement of ambitious programmes on GaAs RISC microprocessors and GaAs-based supercomputers.

The pace of silicon miniaturization means that ultimately the challenge to digital circuits from gallium arsenide has to take place at sub-micrometre geometries where silicon technology is already within sight of its basic limit. This is about 0.2 µm minimum feature size, which should be achievable commercially within the next decade. The physical laws which govern the operation of traditional MOSFET devices permit a further reduction to 0.1 µm if liquid-nitrogen temperatures are used, a feature which is already being deployed in recent supercomputer products where a lower temperature improves the silicon mobility. This is a fascinating development, because low-temperature operation strongly favours gallium arsenide for which a wide range of high-performance device structures exist, and for which devices have already been demonstrated on space scales below 0.1 µm.

Michael Shur's useful and timely book is the first thorough introduction to the understanding of classical GaAs devices and circuits, with an emphasis on very small devices. The material should also provide an excellent background for researchers working on the exciting new range of quantum devices fabricated with gallium arsenide. The strength of Shur's approach comes from his involvement in the large US programme on sub-micrometre electronics, which has explored the fundamental limits of conventional device structures as well as building a rich variety of novel devices. A wealth of references will ensure that this remains a valuable source book for many years. The breakdown of conventional device models at small dimensions is significant at larger scales in gallium arsenide than in silicon, and this area is particularly well discussed, especially classical ballistic electron transport.

As well as describing the wide-ranging GaAs logic families which are of immediate commercial interest, Shur has produced a good introduction to the theory and modelling of modulation doped field-effect transistors and other devices such as the heterojunction bipolar transistor which are based on GaAs-(GaAs)_x(AlAs)_{1-x} layered heterostructures. The potential strength of gallium arsenide is made very clear in these areas, where the flexibility for materials design and innovation far exceeds the present capabilities of silicon. □

John R. Barker is a member of the Nanoelectronics Research Centre and a Professor in the Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, UK.

Diet, brain and behaviour

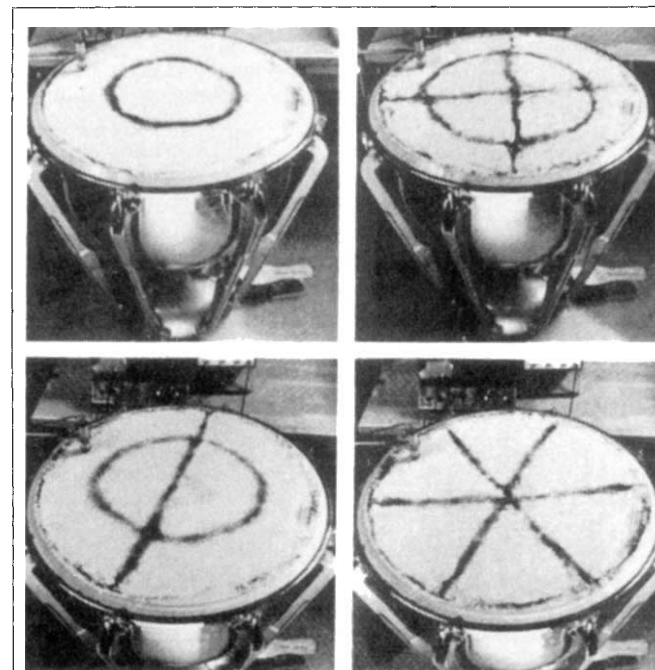
Kevin Connolly

The Effects of Undernutrition on Children's Behavior. By David E. Barrett and Deborah A. Frank. *Gordon & Breach*: 1987. Pp.347. \$88.

THERE is little new about the famines which beset the world, except perhaps that more of us see their sickening manifestations on television. Starvation is not too difficult to identify but malnutrition, despite being one of the most widespread threats to the lives and wellbeing of children, is not easy to define and measure. Broadly speaking there are two approaches — to record dietary intake or to measure physical status. The first of these may seem simple but is all but impossible practically; moreover, there are no agreed reference levels and there are certainly individual and population differences in requirements. The measurement of physical status, particularly weight for height, is the most widely used index, but satisfactory reference standards have only been available since 1978. Despite the difficulties of definition and measurement it is plain that malnutrition is a massive threat to children in the Third World and a considerable threat to a large number in the rich industrial countries of the West.

If severe enough, undernutrition leads to death, before which it impairs growth and increases the individual's susceptibility to disease. Whether it has consequences for behaviour and mental function is obviously an important question. Reasonably, it was assumed that, if malnutrition affects brain growth, then it may in turn have psychological consequences. Important work during the 1960s and 1970s showed that, depending on the timing, duration and severity of the insult, malnutrition did indeed affect the growth of brain in laboratory animals. However we now know that most of the physical changes are reversible, and that the plasticity of the nervous system extends well beyond the relatively short period of early brain development.

Barrett and Frank provide a wide-ranging summary and critique of problems associated with investigating the effects of protein-energy malnutrition on behaviour and psychological function. The book contains five chapters, each of which is divided up by sub-headings. The overall structure and organization is such that a deceptively large amount of information is succinctly presented. The first two chapters deal with fundamental methodological matters — the first with defining and assessing malnutrition, the second with issues related to behaviour (the measure-



Good vibrations — dark powder, scattered on the vibrating head of a kettle drum, migrates to the nodes of the oscillations at which there is no movement. This reveals the various vibrational modes, here excited mechanically, that comprise the characteristic sound of a drum. The harmonic series of such circular, two-dimensional systems is quite different from that of, say, a violin string. The picture is taken from *Fundamentals of Physics*, 3rd edn by D. Halliday and R. Resnick, just published by Wiley. Price is hbk \$53.50, £45.10; pbk \$12.85, £15.95.

ment of behaviour, questions of validity, the design of investigations, inferential limitations and so on). These two chapters are a model of clarity and would make a good source for seminars on methodology.

Chapter 3, the longest in the book, reviews animal and human studies on the effects of general malnutrition on behaviour. From clinical observations on chronically undernourished children it is known that there are correlated psychological effects but it is difficult to tie these directly to malnutrition. Children who are starving or suffering even mild malnutrition also have different social and psychological experiences, so what causes what is difficult to disentangle. Effects there certainly are in laboratory animals, for example on emotional responsiveness, exploration, attentiveness and persistence in problem solving. The findings from work with various groups of children are broadly consistent with animal data. Malnourished infants show impaired attentional processes, reduced social responsiveness, heightened irritability, low activity, diminished affect and so on. The effects of malnutrition on behaviour may be direct, by some physiological route, or indirect through changes in the organism's experience of and interaction with its environment, or by various kinds of covariation between these.

Tucked into the middle of the book is a report on a substantial investigation of chronic undernutrition and child behaviour carried out by the authors on children in a rural part of Guatemala. The ideas behind the study, and details of its design, methods and findings are presented. The findings are consistent with other work — malnourished children tend to be withdrawn and to develop patterns of behaviour that are not conducive to acquiring

social skills; infants tend to be lethargic and passive, lacking attentional capacity and having a low tolerance of stressful stimuli.

The implications, theoretical and practical, are drawn out in the final chapter and they are considerable. That there are behavioural changes associated with chronic malnutrition is clear enough but the causal nexus remains to be disentangled and understood. For example, are the problems identified at school age and later a direct result of organic changes or are they a consequence of disturbed and distorted patterns of interaction in infancy? From the work described and analysed in the book it is plain enough that much remains to be sorted out. For example, we do not know whether there is a critical period in brain development when malnutrition combined with experiential impoverishment will have permanent detrimental effects on the structure and function of the central nervous system. Possible differences in the behavioural effects of acute and chronic malnutrition remain to be explored, and the relative vulnerability of different psychological functions has hardly been considered. To what extent the behavioural consequences of malnutrition can be prevented, ameliorated or reversed all require investigation. Also, behavioural rehabilitation programmes need to be devised and their efficacy investigated.

In his *Advice to a Young Scientist*, Sir Peter Medawar emphasized the need to study important problems. From a scientific and humanitarian point of view there can be few more important problems than the subject of this excellent book. □

Kevin Connolly is Professor of Psychology and Dean of the Faculty of Science, University of Sheffield, Sheffield S10 2TN, UK.

Neutrality and acidity

Peter Brimblecombe

Acid Rain: Rhetoric and Reality. By Chris C. Park. Methuen: 1987. Pp.272. £30, \$39.95. (Methuen Academic has now been incorporated into Routledge.)

RHETORIC and reality are imaginative concepts to apply to the 'acid rain debate'. The words roll off the tongue with such ease that one might wonder if the title of this book isn't itself mere rhetoric. The reader does not have to go far into the text, however, to discover that Chris Park has been careful to give us a broad and informative account of the subject.

Anyone who writes about acid rain risks being accused of bias; the arguments are such that even the most neutral author will be forced to take stands which will ally him with one school of thought or another. It is an indication of Park's struggle for a position of thoughtful neutrality that most informed readers will be able to find points of disagreement with him.

His first difficulty was that of defining acid rain. The choice of definition is frequently driven by a need to explain the effects attributed to acid rain more than by a desire to define some physically observable entity. As a result, the chemical description of acid rain as a pollutant has broadened over the years. Park avoids the widest of definitions by seeing acid rain as derived largely from anthropogenic SO₂ and NO_x emissions. This may explain his cursory treatment of ozone damage to forests and his subtle shifts between questions of urban and of rural pollution. True, reading carefully, it is clear when the author is writing of acid rain and when he is considering urban air pollution, but for many the two may seem synonymous.

The book is intended for the general reader, and source materials have been gleaned from *The Times* almost as often as from the scientific literature; the result is a sense of immediacy. References to scientists tend to be personal, creating an image of them as real people grappling with a complex problem. By contrast, Greenpeace, Friends of the Earth and, more especially, the Central Electricity Generating Board, emerge as monoliths.

The book centres on Western Europe and North America, so we only catch glimpses of the enormous technical and administrative problems caused by the acid-rain problem in Eastern Europe. The author treats the slow emergence of scientific understanding, public interest and legislation very well. The complexities of air and soil chemistry are mentioned rather than discussed in detail, but Park takes pains to show that the public airing

of those complexities by scientists is often more than mere rhetoric designed to impede progress in resolving the problem. The chapter on technological solutions was one of my favourites, no doubt because the sociological and legislative solutions are so much more complicated.

There is such a huge volume of literature on the issues surrounding acid rain that an author has difficulty in deciding where to stop. For the general reader, Park may have erred on the generous side. Conversely, although much of the

material will be well known to specialists, there is plenty to rekindle their imaginations. For example Park reminded me relatively early in the book that atmospheric acidity could arise from a decline in the alkalinity of the atmosphere, a result of removing alkaline fly ash particles from large coal-burning sources. Altogether this book is a praiseworthy effort making 'reality' accessible to a wide audience. □

Peter Brimblecombe is a Lecturer in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Fooling predators

J.L. Cloudsley-Thompson

The Biology of the Reptilia, Vol. 16 Ecology B: Defense and Life History. Edited by Carl Gans and Raymond B. Huey. Alan R. Liss: 1988. Pp.659. \$74.50, £66.

How can a brightly coloured coral snake serve as a batesian model for less-harmful species if its bite is lethal? This paradox has led several investigators to suggest that mildly poisonous rear-fanged species with coral-snake patterns may be the models, while non-venomous and front-fanged species are respectively batesian and müllerian mimics. The name mertenian mimicry has been applied to this hypothesis.

According to F. Harvey Pough, who discusses mimicry and related phenomena in this the latest volume in the series *Biology of the Reptilia*, uncertainties about the probability of envenomation of a predator by a snake, and the toxicity of different venoms, preclude complete evaluation of the hypothesis. Colubrid snakes may be weak müllerian mimics of venomous coral snakes, but they are unlikely to be the species on which the mimicry complex is based. Indeed, coral snakes may well be avoided mainly because they are unpalatable — scavenging birds in Costa Rica, which quickly consume other snakes killed on the roads, apparently leave dead coral snakes undisturbed. Because of the improbability of observing events such as encounters between predators and models or their mimics in nature, the evaluation of alleged examples of mimicry has to be based on indirect evidence.

The arguments outlined above are typical of many to be found in this volume, which is dedicated primarily to topics of ecology and behaviour. The first chapter, by Harry W. Greene, consists of a review of anti-predator mechanisms. Although reptiles frequently prey on invertebrates, the reverse is exceptional. Nevertheless, large centipedes, spiders and scorpions do sometimes eat reptiles and other verte-

brates; turtles are occasionally the prey of octopuses, while crabs constitute formidable predators of their hatchlings. Not much is known of the defence mechanisms of turtles or crocodiles, and morphological defences in general have been little studied — apart from epidermal sound production and hissing in some snakes, the protective shells of tortoises and turtles, and the erection of frills by the bearded dragon *Amphibolurus barbatus*. Such structural features challenge the evolutionary biologist because they often reflect the functional demands of multiple biological roles.

The examples cited above show that the study of reptiles has wider implications than the non-herpetologist might realize. Nor is there any lack of 'corroborative detail' in the book, even if the narrative were to be thought of as 'bald and unconvincing'. For instance, E. N. Arnold devotes an entire chapter to caudal autotomy, while Richard Shine contributes another on parental care. The proportion of species showing this is lower among reptiles than among amphibians or freshwater teleost fishes, a difference which may reflect the absence of male parental care in squamates.

The second half of this large book consists of four multi-author chapters which outline methods for the study of reptile populations, evolution in turtles, life-history patterns in squamate reptiles, and the physiological ecology of reptilian eggs and embryos. Not only do they indicate the enormous amount of field work that has been undertaken in recent years but, even more interesting, they draw attention to the numerous gaps that still exist in our knowledge of reptile biology. □

J.L. Cloudsley-Thompson, Department of Biology (Medawar Building), University College London, Gower Street, London WC1E 6BT, UK, is Emeritus Professor of Zoology in the University of London.

● Recently published by Cambridge University Press in their Studies in Ecology series is *Community Ecology and Salamander Guilds* by Nelson G. Hairston, Sr. The salamander is a good subject of research because it exists in stable populations, and is used here as an example to illustrate the theories of community ecology. Price is £27.50, \$44.50.

A 3,800-million-year isotopic record of life from carbon in sedimentary rocks

Manfred Schidlowski

Max-Planck-Institut für Chemie (Otto-Hahn-Institut), D-6500 Mainz, FRG

An increased ratio of ^{12}C to ^{13}C , an indicator of the principal carbon-fixing reaction of photosynthesis, is found in sedimentary organic matter dating back to almost four thousand million years ago—a sign of prolific microbial life not long after the Earth's formation. Partial biological control of the terrestrial carbon cycle must have been established very early and was in full operation when the oldest sediments were formed.

CARBON is the key element of life and consists principally of a mixture of two stable isotopes, one light (^{12}C) and one heavy (^{13}C); a third, short-lived radioactive nuclide, ^{14}C , occurs only in trace amounts. Pioneering studies^{1,2} established that conversion of inorganic carbon (mostly carbon dioxide) to living matter by biochemical pathways, detailed in Box 1, entails a conspicuous fractionation of the two stable isotopes. Subsequent work^{3–6} has confirmed that all common photosynthetic pathways discriminate against ^{13}C , mainly as a result of a kinetic effect in the first carbon-fixing carboxylation reaction. Consequently, biogenic substances produced by the action of living organisms display a marked preference for light carbon and the heavy carbon is retained preferentially in the surface reservoir of oxidized carbon, mostly as dissolved bicarbonate in seawater (Fig. 1). From our present record of sedimentary carbon isotopes, we can be reasonably confident that biologically mediated isotope fractionation has persisted on Earth quite uniformly for 3.5–3.8 Gyr (refs 7–12). Here we review briefly the evidence for this conclusion and discuss the geochemical implications of an early biological modulation of the terrestrial carbon cycle. The variation of the organic carbon/inorganic carbonate ratio in sedimentary rocks is interpreted here in terms of nutrient control by phosphate but has alternatively been linked to the general oxidation state of the Earth's crust.

Biochemistry of carbon isotope fractionation

Kinetic isotope fractionations in favour of the lighter carbon isotope are primarily imposed on two steps in the primary metabolism of autotrophic organisms that fix atmospheric CO_2 ^{13,14}. These are the uptake and diffusion of external carbon dioxide to the photosynthetic or chemosynthetic reaction centres, and the first irreversible enzymatic carboxylation reaction by which carbon dioxide is incorporated into the carboxyl group of a carboxylic acid. Equilibrium fractionations that also discriminate in favour of a specific isotope are important in alternative pathways that use bicarbonate and are supposedly essential in inter- and intramolecular isotope exchange among isotopically distinct metabolites such as lipids, carbohydrates and proteins¹⁵. Carbon isotope composition differences are usually expressed in terms of $\delta^{13}\text{C}$ values as described in Box 2.

As a whole, the Earth's standing biomass is enriched in isotopically light carbon relative to the inorganic carbon pool of the ocean-atmosphere system, as evidenced by its principal $\delta^{13}\text{C}_{\text{org}}$ spread between -20 and -30% (Fig. 2). Gross fractionations may vary over a considerable range according to which step in the assimilatory pathway (cf. Fig. 3) is rate-controlling, the carboxylation option used (Box 1), and how effectively the isotope fractionations are counteracted by those of the reverse (dissimilatory) processes; the highly variable interactions of these processes express themselves differently in different types

of plants and autotrophic microorganisms (Fig. 2). Most carbon transfer from the non-living to the living realm proceeds by the ribulose 1,5-bisphosphate (RuBP) carboxylase reaction (Box 1, reaction 1), which feeds CO_2 directly into the Calvin cycle, the principal biochemical mechanism for reducing CO_2 to carbohydrate. Because the carboxylation product generated in this process is a compound with a 3-carbon skeleton (phosphoglycerate), the corresponding pathway has been termed C3 photosynthesis. This sequence of carbon assimilation is operated by all green plants, eukaryotic algae and most photosynthetic bacteria; green plants relying on it entirely are termed C3 plants¹⁶. A quantitatively less important carboxylation reaction for fixing CO_2 is found in the so-called C4 plants that use phosphoenolpyruvate (PEP) carboxylase (Box 1, reaction 2) to yield a C4 compound (oxaloacetate). This process entails a minor fractionation of between -2 and -3% relative to the bicarbonate ion that serves

1. Principal CO_2 -fixing carboxylation reactions responsible for the buildup of terrestrial biomass

Reduction of CO_2 primarily gives rise to C3 compounds (with 3-carbon skeletons such as phosphoglycerate and pyruvate), C4 compounds (oxaloacetate) and C2 compounds (acetate, acetyl coenzyme A). The quantitatively most important enzymatic carboxylation is the RuBP carboxylase reaction (1) of the Calvin cycle that forms the initial carbon-fixing reaction in C3 photosynthesis.

Abridged from Schidlowski *et al.*¹⁰.

- (1) $\text{CO}_2 + \text{ribulose-1,5-bisphosphate} \rightarrow \text{phosphoglycerate}$
Operates in: C3 plants*, algae*, cyanobacteria*, purple photosynthetic bacteria (Chromatiaceae)*, chemoautotrophic bacteria.
- (2) $\text{CO}_2/\text{HCO}_3^- + \text{phosphoenolpyruvate/pyruvate} \rightarrow \text{oxaloacetate}$
Operates in: C4 plants*, CAM plants*, anaerobic and facultatively anaerobic bacteria. C4 and CAM plants combine this carboxylation with reaction (1).
- (3) $\text{CO}_2 + \text{CO}_2 \rightarrow \text{acetate/acetyl coenzyme A}$
Operated in: Green photosynthetic bacteria (Chlorobiaceae)* [primary CO_2 fixation is by succinyl coenzyme A and α -ketoglutarate]; anaerobic bacteria, methanogenic bacteria*†.
- (4) $\text{CO}_2 + \text{acetyl coenzyme A} \rightarrow \text{phosphoenolpyruvate/pyruvate}$
Operates in: Green photosynthetic bacteria (Chlorobiaceae)*, combined with reactions (3) and (2); autotrophic sulphate-reducing bacteria, methanogenic bacteria*†.

* Groups of organisms for which carbon isotope fractionations are known (see also Fig. 2).

† Primary CO_2 fixation probably by C1-acceptors; details of the assimilatory pathway of methanogens are poorly known as yet, but the presence of both α -ketoglutarate and pyruvate synthases suggests that reactions (3) and (4) are involved.

2. Carbon isotope composition

Differences in the carbon isotope composition of substances are conventionally expressed as $\delta^{13}\text{C}$ values which give the permil deviation of the $^{13}\text{C}/^{12}\text{C}$ ratio of a sample relative to that of the conventional standard (PDB) that defines zero permil on the δ -scale:

$$\delta^{13}\text{C} = \left[\frac{(^{13}\text{C}/^{12}\text{C})_{\text{sample}}}{(^{13}\text{C}/^{12}\text{C})_{\text{standard}}} - 1 \right] \times 1,000 (\text{‰}, \text{PDB}).$$

A positive value for $\delta^{13}\text{C}$ indicates a higher $^{13}\text{C}/^{12}\text{C}$ ratio (and hence ^{13}C enrichment) in the sample relative to the standard, and a negative value a lower ratio with corresponding ^{13}C depletion. The name of the standard (PDB) is a modified acronym of Peedee belemnite, referring to the carbonate skeleton of a fossil cephalopod (*Belemnitella americana*) from the Cretaceous Peedee Formation which has $^{12}\text{C}/^{13}\text{C} = 88.99$.

as feeder species¹⁷⁻¹⁹ and, coupled with the use of isotopically heavy bicarbonate, is responsible for the relative enrichment of ^{13}C in C4 plants (Fig. 2). The C4 pathway is a relatively late achievement in the evolution of flowering plants and is based on a specific (compartmented) leaf anatomy designed to counteract the unfavourable effects of photorespiration in C3 plants, where part of the newly generated biomass is lost to the atmosphere as CO_2 . Succulent plants with crassulacean acid metabolism (CAM) can fix carbon by both the RuBP and PEP carboxylase reactions albeit at different times, so their compositions span the combined spreads of C3 and C4 plants (Fig. 2). Green photosynthetic sulphur bacteria (Chlorobiaceae) have a consistently low fractionation²¹⁻²³ approaching the range of C4 plants (Fig. 2); they rely on a quantitatively less important ferredoxin-linked carboxylation (Box 1, reaction 3) within the reductive carboxylic acid cycle, a pathway unique to photosynthetic prokaryotes²⁰. Diffusion-limited autotrophic pathways preferentially operated by aquatic plants and microorganisms²⁴⁻²⁷ entail small fractionations (-4% and less) and make virtually no impact on the global isotope budget because they are confined to minor side stages of the carbon cycle.

Figure 3 is a synopsis of the principal isotope-discriminating steps in autotrophic (principally photosynthetic) carbon fixation. In sum, these express themselves as a sizeable negative shift of the $\delta^{13}\text{C}$ values of biosynthesized matter relative to the feeder substrate (mostly atmospheric CO_2). The kinetic isotope effect in the initial diffusion step has been shown to be small, at the limit being equal to the value for CO_2 diffusion in air (-4.4%) (refs 13, 14). In contrast, fractionations in subsequent enzymatic carboxylation reactions are usually much larger, though distinctly variable, being mainly in the range of -20 to -40% for the quantitatively most important RuBP carboxylase reaction^{13,14,28}. Thus the contemporary biomass bears, in essence, the isotopic signature of C3 photosynthesis characterized by this sizeable isotope effect exercised by the key enzyme of the Calvin cycle that channels most of the carbon transfer to the living realm, with a resulting $\delta^{13}\text{C}_{\text{org}}$ range of about $-26 \pm 7\%$ (Fig. 2).

Preservation of past isotope fractionations

From the exchange reservoir on the Earth's surface, a steady flux of both organic carbon (C_{org}) and carbonate carbon (C_{carb}) enters newly formed sediments. In this way, a formidable carbon reservoir of $\sim 10^{22}$ g has accumulated over the ages within the Earth's crust (Fig. 1), about one-fifth of which is organic carbon, mostly in the form of kerogen, the highly polymerized, acid-insoluble end-product of the diagenetic (post-depositional) reconstitution of dead organic matter. Kerogen and its derivative graphite make up 0.5–0.6% of the total mass of sedimentary rocks from the beginning of the geological record^{30,31}. Moreover, both inorganic and organic sedimentary carbon have been shown to retain the isotopic compositions of their progenitors

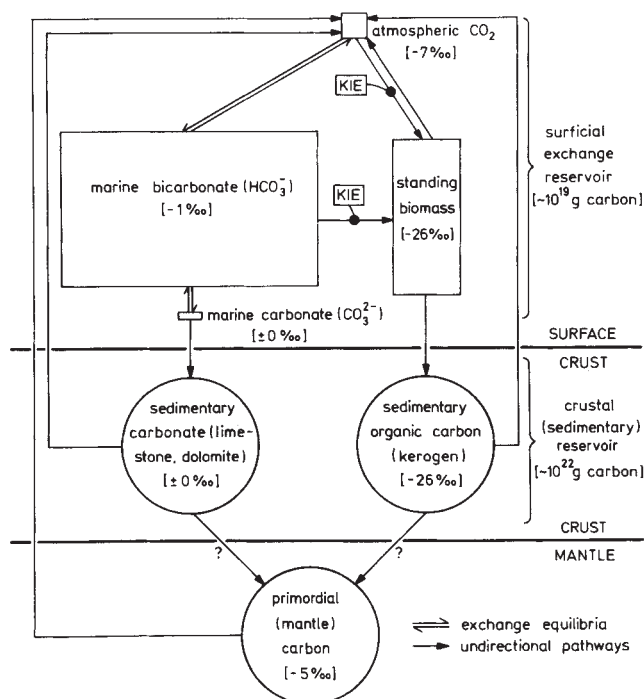


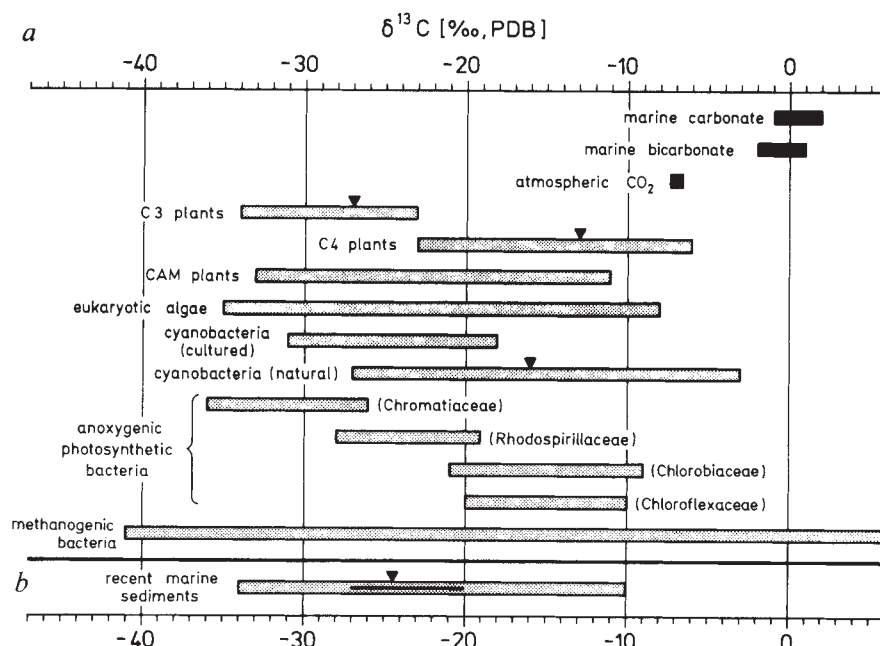
Fig. 1 Flow diagram of the carbon cycle showing surficial, crustal and mantle reservoirs. Proportions of surficial compartments (rectangular units) are approximately to scale; $\delta^{13}\text{C}$ averages of individual reservoirs are given in brackets (see Box 2). Note that the standing biomass is about one order of magnitude, and the atmospheric carbon dioxide reservoir about two orders of magnitude smaller than the ocean's burden of dissolved bicarbonate. Fluxes of inorganic carbon into the biosphere are subject to a kinetic isotope effect (KIE) responsible for a preferential accumulation of ^{12}C in organic matter, with the heavy complement (^{13}C) basically retained in marine bicarbonate. A continuous flux of carbonate and organic carbon from the surface enters newly formed sediments, propagating the isotopic difference between the two carbon species into the rock section of the cycle (circled boxes). Here, this disparity is basically 'frozen' until the host sediments undergo profound transformations by metamorphism and remelting. In the fullness of time, crustal carbon may be recycled and re-enter the surficial reservoir as CO_2 .

in the surficial environment. Commonly, carbonate rocks preserve the isotopic composition of their parent muds within $\pm 1\%$ of the original value which was, in turn, inherited from a bicarbonate precursor with a small shift of $+1\%$ due to the equilibrium fractionation between HCO_3^- and CO_3^{2-} (ref. 32). Accordingly, sedimentary carbonates monitor, to a good approximation, the isotopic composition of marine bicarbonate through time. On the other hand, there may be a ^{13}C enrichment of 2–3% in mature kerogens, resulting from isotopic changes during burial and diagenesis of organic matter, mainly through preferential removal of ^{13}C -depleted volatiles that subsequently accumulate as oil and gas during maturation^{9,10,15,33-36}. As a whole, however, the diagenetically stabilized kerogen fraction of unmetamorphosed sediments basically preserves the kinetic isotope effect involved in the formation of the biological source materials, so that the isotopic signature of autotrophic carbon fixation can be traced back into the geological past.

The isotope record

Because organic carbon and carbonate are abundantly preserved in sedimentary rocks over a span of 3.8 Gyr, and the secondary isotope effects linked to their post-depositional alteration are reasonably well understood, the collection of a comprehensive body of carbon isotope data for the whole record seems a promising way to identify the time when autotrophic carbon

Fig. 2 *a*, Isotopic composition of major groups of higher plants and autotrophic microorganisms¹² compared to that of the oxidized carbon compounds (CO_2 , HCO_3^- , CO_3^{2-}) of the surficial environment. Triangles indicate the approximate mean. The consistently negative $\delta^{13}\text{C}$ values of organic matter imply a depletion in ^{13}C (or, conversely, an enrichment in ^{12}C) relative to the oxidized carbon source species. The positive extremes for methanogens were obtained in culture experiments under specific conditions that are irrelevant for natural communities⁸⁷. *b*, Isotope spread for organic carbon constituents of recent marine sediments⁸⁸; the black insert covers >90% of some 1,600 data points summarized in the band.



fixation began. Once photoautotrophic organisms had learned to operate the CO_2 -fixing machine by means of light energy, we can assume that life processes attained partial control of the terrestrial carbon cycle. The development of the water-splitting form of photosynthesis which releases free oxygen as a metabolic by-product was, moreover, the most important single process ever to make its impact on the evolution of the Earth's atmosphere^{7,37}.

During the last two decades, major collections of carbon isotope data from sedimentary rocks have been assembled, starting with the younger (Phanerozoic) section of the record³⁸⁻⁴⁰ and proceeding to progressively older Precambrian rocks^{8,10,15,36,41-45}. As a result, there are ~10,000 measurements of $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ covering the 3.8 Gyr of the known record; this prolific data base is summarized in Fig. 4. Decoding this store of information seems straightforward against the biochemical background summarized above. As sedimentary $\delta^{13}\text{C}_{\text{carb}}$ values fix the values of their bicarbonate precursors with a slight positive shift of only ~1‰, the message conveyed by the carbonate curve is that the isotopic composition of marine bicarbonate has not varied much through geological history. Accepting the well-documented Phanerozoic section⁴⁰ as representative of the whole record, oscillations of $\delta^{13}\text{C}$ values in marine carbonates should fall mainly within $\pm 2.5\%$ around a long-term average of about +0.5‰.

In a similar way, $\delta^{13}\text{C}_{\text{org}}$ values appear to be tethered to a mean somewhere between -24 and -28‰, with a total spread from about -10 to -50‰ and with most data points falling in the range $-26 \pm 7\%$ (Fig. 4). As this spread shows a conspicuous coincidence with the $\delta^{13}\text{C}_{\text{org}}$ ranges of extant primary producers, there is little doubt that the enrichment of ^{12}C in fossil organic matter reflects, for the most part, corresponding enrichments in the biological precursor materials. Accordingly, the sedimentary record of organic carbon conveys a remarkably consistent isotopic signal of autotrophic carbon fixation as from at least 3.5, and maybe as far back as 3.8, Gyr ago. The basic uniformity of this isotopic signature through time is testimony to an extreme degree of evolutionary conservatism in the biochemistry of photoautotrophy as the quantitatively most important process of CO_2 assimilation. It is reasonable to speculate that the main part of the isotopic envelope for fossil organic matter depicted in Fig. 4 is the geochemical manifestation of the ability of one enzyme, namely RuBP carboxylase, to discriminate between light and heavy isotopes of carbon. As the Calvin cycle channels the bulk of the carbon transfer to the living world, it is not

surprising that fossil organic matter should bear the isotopic signature of its principal CO_2 -fixing reaction. Hence, the carbon isotope age curve as a whole constitutes an index line of autotrophic carbon fixation spanning almost 4 Gyr of recorded Earth history.

Noteworthy exceptions are the unusually negative $\delta^{13}\text{C}_{\text{org}}$ values at the lower fringe of the envelope, including two conspicuous offshoots with minima between -47 and -52‰ (Fig. 4). These values come primarily from a number of late Archaean (2.6-2.8 Gyr) kerogens^{10,36,46-48}, but have also been observed in the early Proterozoic (~2.1 Gyr)⁴⁹. As negative $\delta^{13}\text{C}$ values of this range have hitherto only been reported in connection with methane, we must necessarily assume a methane-utilizing step in the formation of the precursors of these kerogens either by direct CH_4 consumption by methylotrophic bacteria, or by common autotrophs using light-isotope CO_2 stemming from the oxidation of CH_4 . It is not surprising that such negative anomalies are found merely in the early Precambrian when, due to lower environmental oxygen pressures, methanogenesis was probably more important than it is today as a key process in the anaerobic decomposition of organic matter.

The oldest record

The only conspicuous discontinuity in the two carbon isotope age curves depicted in Fig. 4 is the break between the oldest terrestrial sediments (3.76 Gyr^{50,51}) from the Archaean terrane

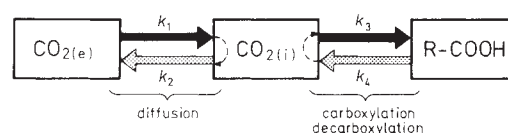
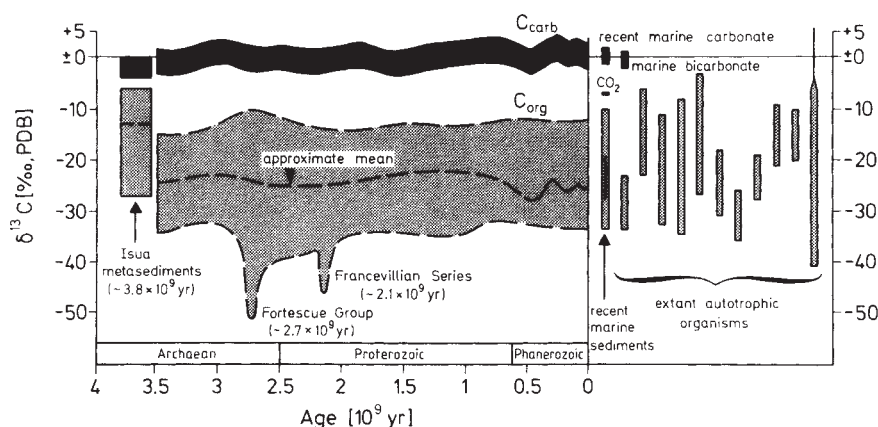


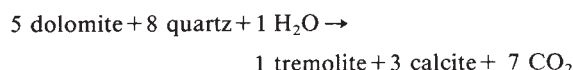
Fig. 3 Summary of the principal isotope-discriminating steps in the primary metabolism of CO_2 -fixing plants and microorganisms (black: assimilatory reactions; stippled: dissimilatory and other reverse processes; k_1 - k_4 , corresponding rate constants). $\text{CO}_{2(e)}$ represents the carbon dioxide of the environmental feeder pool and $\text{CO}_{2(i)}$ the 'internal' CO_2 that has already entered living tissue on its way to the photosynthetically active sites; R-COOH stands for the product of the first CO_2 -fixing enzymatic carboxylation. In sum, these processes result in a preferential accumulation of light carbon (^{12}C) in the synthesized biomass in the right-hand box, the largest single fractionation effect commonly deriving from the isotope-discriminating properties of the carboxylating enzymes operative in the second step.

Fig. 4 Isotopic composition of sedimentary carbonate (C_{carb}) and organic carbon (C_{org}) over 3.8 Gyr of the Earth's history as compared to the isotopic compositions of their progenitor materials in the contemporary environment (marine bicarbonate and biogenic matter). Spreads shown for extant autotrophs are those of Fig. 2, where the top to bottom sequence corresponds to that from left to right shown here. The Archaean and Proterozoic eras indicated on the time axis are conventionally summarized as Precambrian. The envelope for fossil organic carbon is an update of the data base presented by Schidlowski *et al.*¹⁰, comprising the means of some 150 Precambrian kerogen provinces; other parts of the record are based on currently available literature data (see text). Note that the isotope spreads of contemporary autotrophs are transcribed into the record with just the extremes eliminated. Conspicuous negative excursions such as observed in the Fortescue Group (Australia)^{10,36} and the Francevillian of Gabon (Africa)⁴⁹ indicate the involvement of methane in the formation of the kerogen precursors. Both $\delta^{13}C_{org}$ and $\delta^{13}C_{carb}$ values are basically constant over 3.5 Gyr, but have been reset by metamorphism in the Isua suite⁸.



of Isua, West Greenland and the rest of the record⁸⁻¹⁰. Whereas the isotopic composition of organic carbon over the last 3.5 Gyr can be taken as evidence of biological activity, the unexpectedly positive mean for the Isua kerogens and their graphitic derivatives ($\delta^{13}C_{org} = -13.0 \pm 4.9\%$) (ref. 10) raised the question as to whether values in this range constitute equally unequivocal indicators of biogenicity. There is also a minor negative shift of 2-3% in the Isua carbonates (Fig. 4).

We first proposed that the observed isotope changes were caused by the amphibolite-grade metamorphism (cf. Fig. 5) to which the rocks of the Isua Supracrustal Belt had been subjected since they were deposited⁸. A wealth of additional evidence from other metamorphic terranes⁵²⁻⁵⁷ now supports this hypothesis, indicating that shifts such as that observed in the Isua deposits are primarily due to high-temperature reequilibration of isotopes between coexisting sedimentary carbonate and organic carbon (Fig. 5). Specifically, it has been shown that $^{13}C/^{12}C$ exchange between kerogen constituents and the isotopically heavy CO_2 released by the decarbonation reactions that accompany the formation of metamorphic calc silicates, such as tremolite or diopside, for example:



starts during low-grade metamorphism (greenschist facies, 300-450 °C) and increases with increasing metamorphic temperatures⁵⁴. Calculated fractionation factors⁵⁸ indicate that the carbon dioxide liberated by decarbonation reactions is enriched in ^{13}C , shifting the $\delta^{13}C_{carb}$ values of the residual carbonate (calcite) towards more negative values (see corresponding displacement of Isua carbonates in Fig. 4). We may reasonably assume that the CO_2 released forms the principal vehicle for isotope exchange in sediments undergoing metamorphic reconstitution. Isotopic reequilibration between this heavy carbon dioxide and kerogenous materials tends progressively to decrease the magnitude of fractionation between C_{org} and C_{carb} as a function of increasing metamorphic temperature (Fig. 5b), with thermodynamic equilibrium being closely approached in the high-temperature range (≥ 650 °C) of the granulite facies (Fig. 5a).

These findings also allow us to interpret those parts of the record that are clouded by a metamorphic overprint. There is little doubt that the Isua values have been reset by metamorphism, and that the isotopic signature of autotrophic carbon fixation originally extended back at least to Isua times. This, in turn, would suggest that photosynthesis as the second geochemically important CO_2 -fixing process emerged very early in

the Earth's history, with biological processes consequently leaving their mark on the terrestrial carbon cycle as early as almost 4 Gyr ago.

To challenge this interpretation, we would have to postulate an inorganic process operating on a global scale that was capable of mimicking, both in direction and magnitude, the principal enzymatic isotope effect of the photosynthetic pathway with remarkable precision. Potential alternatives can be safely dismissed, as they fail to meet the requirement of a near-perfect isotopic mimicry. Fractionations between oxidized and reduced carbon in Fischer-Tropsch reactions have been shown to be in substantial excess over those of photosynthetic processes, amounting to between -50 and -100‰⁵⁹. Miller-Urey-type spark discharge syntheses conversely entail relatively small isotope effects of -10‰ and less in the case of their bulk reaction products⁶⁰. Though the latter marginally overlap the biological range, they are decidedly not typical of average organic matter and should quickly be detected where they prevail in the organic fraction of unmetamorphosed sediments. Substantial contributions from such sources to the reduced carbon content, notably of the oldest sediments, is excluded by their virtual absence from the record.

It can thus be assumed that autotrophy, particularly photoautotrophy, has been extant on Earth both as a biochemical process and as a geochemical agent for at least 3.8 Gyr. The early appearance of reduced carbon bearing the isotopic signature of autotrophic carbon fixation is, moreover, fully consistent with the paleontological record of apparently autotrophic life⁶¹⁻⁶⁴. Biosedimentary structures attributable to the mat-building activities of prokaryotic microbenthos ('stromatolites') have been traced back to ~3.5 Gyr⁶⁴⁻⁶⁷, but the biogenicity of the presumably oldest cellular microfossils^{68,69} is still disputed⁷⁰⁻⁷². However, with benthic prokaryotes unquestionably present 3.5 Gyr ago, a strong case can be made that their ancestral lines have extended to Isua times, and possibly beyond.

Biogeochemical implications

The near-time invariance on the billion-year scale of the isotopic composition of both sedimentary carbon species has profound implications for their relative proportions in the global geochemical system. Because crustal carbon is derived from the mantle, its bulk isotopic composition is necessarily constrained by the isotope mass balance governing the carbon transfer from this reservoir. Mantle carbon as represented by diamonds, carbonates and carbon constituents of mid-ocean basalts typically has $\delta^{13}C$ spreads between -3 and -8‰ (refs 73-75). If we accept $\delta^{13}C_{prim} = -5\%$ as a reasonable approximation for the isotopic composition of the primary endowment of carbon inherited by

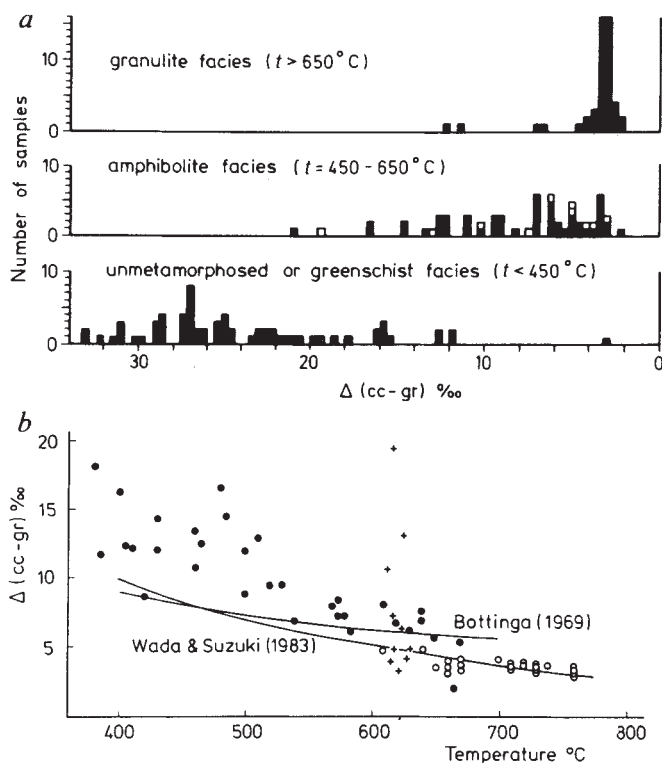


Fig. 5 Evidence that the positive shift in mean $\delta^{13}\text{C}$ for the Isua kerogens results from high-temperature reequilibration of isotopes between coexisting sedimentary carbonate and organic carbon during metamorphic reconstitution of the primary rock. Two ways of presenting the decrease of carbon isotope fractionation between calcite (cc) and graphite (gr) [$\Delta(\text{cc-gr}) = \delta^{13}\text{C}_{\text{cc}} - \delta^{13}\text{C}_{\text{gr}}$] are shown as a function of *a*, metamorphic grade or *b*, increasing temperature (organic carbon tends to occur as graphite at higher metamorphic grades). *a*, In a comparison of unaltered and metamorphic sediments, it is obvious that unmetamorphosed and low-grade rocks mostly display fractionations between 20 and 30‰ (lower data set), whereas $\Delta(\text{cc-gr})$ decreases significantly in amphibolite-grade metasediments and still more in the granulite facies as a result of high-temperature isotopic reequilibration between the two carbon species. Emergence of a narrow peak in the granulite facies indicates that thermodynamic equilibrium is virtually attained at $t \geq 650^\circ\text{C}$. Adapted from ref. 54, with additional data points (open squares) from ref. 57. *b*, Thermodynamically calculated isotope fractionation curve of Bottinga⁵⁸ and an empirical counterpart by Wada and Suzuki⁵⁶ calibrated by dolomite-calcite solvus temperatures; fractionations reported by other investigators from various metamorphic terranes (open circles, ref. 54; filled circles, ref. 55; crosses, ref. 57) scatter around both functions. Markedly discordant values can persist up to 600–650 $^\circ\text{C}$, indicating that isotopic reequilibration between organic carbon and carbonate is not complete even in the upper amphibolite facies.

the Earth from the parent solar nebula, the isotope mass balance

$$\delta^{13}\text{C}_{\text{prim}} = R\delta^{13}\text{C}_{\text{org}} + (1 - R)\delta^{13}\text{C}_{\text{carb}}$$

would give us $R = C_{\text{org}} / (C_{\text{org}} + C_{\text{carb}}) = 0.2$ when substituting the long-term averages of $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ (about -25% and $\pm 0\%$) into the above equation. $R = 0.2$ would, in turn, imply a $C_{\text{org}}/C_{\text{carb}}$ ratio of 0.2/0.8, indicating that C_{org} accounts for 20% and C_{carb} for 80% of total carbon in the sedimentary shell.

As the isotope values providing this information were originally established in the surficial exchange reservoir (Fig. 1) before being encoded in the record, they primarily indicate the above 20%/80% ratio for the surficial compartment of the cycle. If this compartment were flushed at a rapid rate as it is

today (with a statistical residence time of an individual bicarbonate ion in seawater of about 10^5 years⁷⁶), a steady state complying with the mass balance equation above would be established in a geologically short time interval.

Given these constraints, it follows from the long-term near-constancy of the $\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$ age functions that organic carbon had always been about one-fifth of the total carbon in the surficial compartment. Though somewhat puzzling at first, this conclusion is compatible with the principle of biotic plenitude, or fullness, that reflects the intrinsic property of life to proliferate exponentially to limits ultimately set by the availability of critical resources, notably limiting nutrients such as phosphorus and nitrogen. The ancient Earth could have been in such a state of global biotic saturation when prolific microbial ecosystems monopolized the totality of suitable habitats; the surface-dwelling (benthic) varieties have been well preserved in the form of impressive fossil stromatolite carpets since early Archaean times^{64–67}.

With the aquatic realm essentially to themselves, these early microbial (largely prokaryotic) communities probably bloomed to limits determined by nutrient supply, and it is not difficult to imagine ancient habitats saturated with microorganisms to a degree approaching that of the artificially eutrophicated environments of the present world⁷⁷. As the Precambrian continents were virtually barren, there were no land biota to retain rare nutrients like phosphorus, hence the ancient seas were probably eutrophicated by a factor of at least two as compared to the present oceans. In fact, accepting the role of phosphorus as the ultimate determinant for the size of the global biomass^{76,78} implies the existence of a state of global biotic saturation or plenitude from the time when autotrophic life first appeared on Earth. It is, moreover, reasonable to conjecture that the relative constancy of the $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ age functions over almost 4 Gyr is primarily due to the fact that the $C_{\text{org}}/C_{\text{carb}}$ ratio in the atmosphere-ocean-crust system had been held constant throughout the ages by a roughly uniform rate of phosphorus cycling through the exogenic (surficial) compartment.

In this context it is relevant that extant mats of benthic prokaryotes (specifically cyanobacteria) sustain formidable rates of primary production, with maxima around 8–12 g $\text{C}_{\text{org}} \text{m}^{-2}$ per day^{79,80}, implying that microbial communities are among the most productive ecosystems of the present world. Given suitable environmental conditions, there is no reason to believe that respective performances by their ancient precursors were below these standards. Moreover, it is now generally accepted that the Precambrian was the time when prokaryotic ecosystems held dominion over the Earth^{61–63}. If such impressive rates of primary production can be sustained by microbial photoautotrophs, photosynthesis may have gained little in effectiveness and quantitative importance since the first microbial communities established their reign on the ancient Earth. This inference is also consistent with the observation that the average C_{org} content of Precambrian sediments does not seem to differ basically from that of Phanerozoic formations, often with means between 0.5 and 0.6% (refs 30, 31, 81) (including the organic-rich members of the Isua suite³¹).

Empirical evidence for this scenario is still scarce and many details of the picture need elucidation. An alternative to a limiting nutrient control of the $C_{\text{org}}/C_{\text{carb}}$ ratio could, for instance, be that this ratio was preordained to the exogenic system by the general oxidation state of the Earth's outer veneer (and notably of its degassed volatiles)^{76,82}. As the escape of hydrogen from the atmosphere is quantitatively unimportant⁸³, the Earth as a whole may be considered to approximate a closed system, with its overall redox state relatively constant over geological time. Even a fixed oxidation state of the atmosphere-ocean-crust system as a whole should, however, allow certain degrees of freedom for an internal redistribution of redox equivalents within the redox-coupled C–S–Fe–O cycles. Such redistribution is vividly demonstrated in the case of sulphur, where dramatic

oscillations of the $\delta^{34}\text{S}$ age curve of marine sulphate reflect sizeable changes in the sulphide/sulphate ratio over much of the Earth's history^{84,85}. As $C_{\text{org}}/C_{\text{carb}}$ ratios tend to remain remarkably constant under such conditions, it seems reasonable to argue that this ratio was fixed by an independent determinant, most probably phosphate. If phosphorus is a determinant of the maximum size of the Earth's biomass, other uncertainties concern the supply of this element to the oldest habitats of life in the Archaean seas. The productivity of the present oceans is maintained largely by the fluvial phosphate input from the continents, but the scarcity and small size of continental areas, particularly during the early Archaean, suggest a different mode of phosphate cycling at this early stage, with corresponding repercussions on the marine carbon cycle. It is, therefore, most likely that the productivity scheme outlined above applies to late Archaean and Proterozoic times, rather than to the early

Archaean. The complex questions related to the productivity of the Precambrian oceans^{82,86} call for further attention.

In conclusion, photosynthesis must have existed as a biochemical process (primarily operated by microbial ecosystems) for almost 4 billion years, and over the same time have been an important agent in the geochemical transformations of the Earth's surface. This conclusion, based on the currently available geological isotope record, would only be in error if a global process had existed capable of mimicking, in some astounding way, carbon isotope fractionations in photosynthesis.

This article is a contribution to the UNESCO-IUGS sponsored International Geological Correlation Program Project 157 (Early Organic Evolution and Mineral and Energy Resources). Relevant work has been supported in various stages by the Deutsche Forschungsgemeinschaft, lately as part of SFB 73.

- Nier, A. O. & Gulbransen, E. A. *J. Am. chem. Soc.* **61**, 697-698 (1939).
- Murphy, B. F. & Nier, A. O. *Phys. Rev.* **59**, 771-772 (1941).
- Rankama, K. *J. Geol.* **56**, 199-209 (1948).
- Wickmann, F. E. *Geochim. cosmochim. Acta* **2**, 243-254 (1952).
- Craig, H. *Geochim. cosmochim. Acta* **3**, 53-92 (1953).
- Park, R. & Epstein, S. *Geochim. cosmochim. Acta* **21**, 110-126 (1960).
- Junge, C. E., Schidlowski, M., Eichmann, R. & Pietrek, H. *J. geophys. Res.* **80**, 4542-4552 (1975).
- Schidlowski, M., Appel, P. W. U., Eichmann, R. & Junge, C. E. *Geochim. cosmochim. Acta* **43**, 189-199 (1979).
- Schidlowski, M. in *Cosmochemistry and the Origin of Life* (ed. Ponnampetura, C.) 277-322 (Reidel, Dordrecht, 1983).
- Schidlowski, M., Hayes, J. M. & Kaplan, I. R. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 149-186 (Princeton University Press, 1983).
- Schidlowski, M. *Adv. Space Res.* **4**(12), 183-193 (1984).
- Schidlowski, M. *A. Rev. Earth planet. Sci.* **15**, 47-72 (1987).
- Vogel, J. C. *Sitzungsber. Heidelb. Akad. Wiss. Math. Naturwiss. Kl.* **3**, 111-135 (1980).
- O'Leary, M. H. *Phytochemistry* **20**, 553-567 (1981).
- Galimov, E. M. *The Biological Fractionation of Isotopes* (Academic, Orlando, Florida, 1985).
- Walker, D. *Energy, Plants and Man* (Packard, Chichester, 1979).
- Reibach, P. H. & Benedict, C. R. *Pl. Physiol.* **59**, 564-568 (1977).
- Schmidt, H. L., Winkler, F. J., Latzko, E. & Wirth, E. *Israel J. Chem.* **17**, 223-224 (1978).
- O'Leary, M. H. *A. Rev. Pl. Physiol.* **33**, 297-315 (1982).
- Buchanan, B. B. in *Encyclopedia of Plant Physiology* 6 (eds Gibbs, M. & Latzko, E.) 416-424 (Springer, Berlin, 1979).
- Barghoorn, E. S., Knoll, A. H., Dembicki, H. & Meinschein, W. G. *Geochim. cosmochim. Acta* **41**, 425-430 (1977).
- Quandt, L., Gottschalk, G., Ziegler, H. & Stichler, W. *FEMS Mikrobiol. Lett.* **1**, 125-128 (1977).
- Sirevåg, R., Buchanan, B. B., Berry, J. A. & Troughton, J. H. *Arch. Mikrobiol.* **112**, 35-38 (1977).
- Benedict, C. R., Wong, W. W. L. & Wong, J. H. H. *Pl. Physiol.* **65**, 512-517 (1980).
- LaZerte, B. D. & Szalados, J. E. *Limnol. Oceanogr.* **27**, 413-418 (1982).
- Schidlowski, M., Matzigkeit, U. & Krumbein, W. E. *Naturwissenschaften* **71**, 303-308 (1984).
- Schidlowski, M., Matzigkeit, U., Mook, W. G. & Krumbein, W. E. in *Hypersaline Ecosystems: The Gavish Sabkha* (Ecol. Stud. **53**) (eds Friedman, G. M. & Krumbein, W. E.) 381-401 (Springer, Berlin, 1985).
- Winkler, F. J., Kexel, H., Kranz, C. & Schmidt, H. L. in *Stable Isotopes* (Anal. Chem. Symp. Ser. **11**) (eds Schmidt, H. L., Förstel, H. & Heinzinger, K.) 83-89 (Elsevier, Amsterdam, 1982).
- Durand, B. (ed.) *Kerogen—Insoluble Organic Matter from Sedimentary Rocks* (Editions Technip, Paris, 1980).
- Ronov, A. B. *Osadochnaya Obolochka Zemli* (Earth's Sedimentary Shell: 20th Vernadsky Lecture (Izdatel'stvo Nauka, Moscow, 1980).
- Schidlowski, M. in *Mineral Deposits and the Evolution of the Biosphere* (eds Holland, H. D. & Schidlowski, M.) 103-122 (Springer, Berlin, 1982).
- Vogel, J. C. in *Summer Course on Nuclear Geology, Varenna 1960* (eds Houtermans, F. G., Picciotto, E. E. & Tongioli, E.) 216-221 (Lischi, Pisa, 1961).
- Sackett, W. M., Nakaparksin, S. & Dalrymple, D. in *Advances in Organic Geochemistry 1966* (eds Hobson, G. D. & Speers, G. C.) 37-53 (Pergamon, Oxford, 1968).
- Chung, H. M. & Sackett, W. M., *Geochim. cosmochim. Acta* **43**, 1979-1988 (1979).
- Peters, K. E., Rohrbach, B. G. & Kaplan, I. R. *Bull. Am. Ass. Petrol. Geol.* **65**, 501-508 (1981).
- Hayes, J. M., Kaplan, I. R. & Wedeking, K. W. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 93-134 (Princeton University Press, 1983).
- Schidlowski, M. *J. Geol. Soc. Lond.* **141**, 243-250 (1984).
- Keith, M. L. & Weber, J. M. *Geochim. cosmochim. Acta* **28**, 1787-1816 (1964).
- Degens, E. T. in *Organic Geochemistry* (eds Eglinton, G. & Murphy, M. T. J.) 304-329 (Springer, Berlin, 1969).
- Veizer, J., Holser, W. T. & Wilgus, C. K. *Geochim. cosmochim. Acta* **44**, 579-587 (1980).
- Oehler, D. Z., Schopf, J. W. & Kvenvolden, K. A. *Science* **175**, 1246-1248 (1972).
- Eichmann, R. & Schidlowski, M. *Geochim. cosmochim. Acta* **39**, 585-595 (1975).
- Schidlowski, M., Eichmann, R. & Junge, C. E. *Precamb. Res.* **2**, 1-69 (1975).
- Schidlowski, M., Eichmann, R. & Fiebigler, W. *Neues Jb. Miner. Mh.* **1976**, 344-353 (1976).
- Veizer, J. & Hoefs, J. *Geochim. cosmochim. Acta* **40**, 1387-1395 (1976).
- Hayes, J. M. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 291-301 (Princeton University Press, 1983).
- Schoell, M. & Wellmer, F. W. *Nature* **290**, 696-699 (1981).
- Strauss, H. *Geochim. cosmochim. Acta* **50**, 2653-2662 (1986).
- Weber, F., Schidlowski, M., Arneith, J. D. & Gauthier-Lafaye, F. *Terra Cognita* **3**, 220 (1983).
- Moorbath, S., O'Nions, R. K., Pankhurst, R. J., Gale, N. H. & McGregor, V. R. *Nature phys. Sci.* **240**, 78-82 (1972).
- Moorbath, S., O'Nions, R. K. & Pankhurst, R. J. *Nature* **245**, 138-139 (1973).
- McKirdy, D. M. & Powell, T. S. *Geology* **2**, 591-595 (1974).
- Hoefs, J. & Frey, M. *Geochim. cosmochim. Acta* **40**, 945-951 (1976).
- Valley, J. W. & O'Neil, J. R. *Geochim. cosmochim. Acta* **45**, 411-419 (1981).
- Kreulen, R. & van Beek, P. C. J. *Geochim. cosmochim. Acta* **47**, 1527-1530 (1983).
- Wada, H. & Suzuki, K. *Geochim. cosmochim. Acta* **47**, 697-706 (1983).
- Arneith, J. D., Schidlowski, M., Sarbas, B., Goerg, U. & Amstutz, G. C. *Geochim. cosmochim. Acta* **49**, 1553-1560 (1985).
- Bottinga, Y. *Geochim. cosmochim. Acta* **33**, 49-64 (1969).
- Lancet, M. S. & Anders, E. *Science* **170**, 980-982 (1970).
- Chang, S., DesMarais, D., Mack, R., Miller, S. R. & Strathearn, G. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 53-92 (Princeton University Press, 1983).
- Cloud, P. E. *Paleobiology* **2**, 351-387 (1976).
- Awramik, S. M. in *Mineral Deposits and the Evolution of the Biosphere* (eds Holland, H. D. & Schidlowski, M.) 67-81 (Springer, Berlin, 1982).
- Schopf, J. W. (ed.) *Earth's Earliest Biosphere: Its Origin and Evolution* (Princeton University Press, 1983).
- Walter, M. R. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 187-213 (Princeton University Press, 1983).
- Dunlop, J. S. R., Muir, M. D., Milne, V. A. & Groves, D. I. *Nature* **274**, 676-678 (1978).
- Lowe, D. R. *Nature* **284**, 441-443 (1980).
- Walter, M. R., Buick, R. & Dunlop, J. S. R. *Nature* **284**, 443-445 (1980).
- Pflug, H. D. & Jaeschke-Boyer, H. *Nature* **280**, 483-486 (1979).
- Awramik, S. M., Schopf, J. W. & Walter, M. R. in *Developments and Interactions of the Precambrian Atmosphere, Lithosphere and Biosphere* (eds Nagy, B., Weber, R., Guerrero, J. C. & Schidlowski, M.) 249-266 (Elsevier, Amsterdam, 1983).
- Bridgwater, D. *et al. Nature* **289**, 51-53 (1981).
- Roedder, E. *Nature* **293**, 459-462 (1981).
- Buick, R., *Precamb. Res.* **24**, 157-172 (1984).
- Deines, P. & Gold, D. P. *Geochim. cosmochim. Acta* **37**, 1709-1733 (1973).
- DesMarais, D. J. & Moore, J. G. *Earth planet. Sci. Lett.* **69**, 43-57 (1984).
- Mattey, D. P., Carr, R. H., Wright, I. P. & Pillinger, C. T. *Earth planet. Sci. Lett.* **70**, 196-206 (1984).
- Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, New York, 1978).
- Stanley, S. M. *The New Evolutionary Timeable* (Basic Books, New York, 1981).
- Broecker, W. S. in *Carbon and the Biosphere* (eds Woodwell, C. M. & Pecan, E. V.) 32-50 (US Atomic Energy Commission, Washington, D.C., 1973).
- Krumbein, W. E. & Cohen, Y. in *Fossil Algae* (ed. Flügel, E.) 37-56 (Springer, Berlin, 1977).
- Cohen, Y., Aizenshtat, Z., Stoler, A. & Jørgensen, B. B. in *Biogeochemistry of Ancient and Modern Environments* (eds Ralph, J. B., Trudinger, P. A. & Walter, M. R.) 167-172 (Springer, Berlin, 1980).
- Cameron, E. M. & Garrels, R. M. *Chem. Geol.* **28**, 181-197 (1980).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton University Press, 1984).
- Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
- Schidlowski, M., Junge, C. E. & Pietrek, H. *J. geophys. Res.* **82**, 2557-2565 (1977).
- Claypool, G. E., Holser, W. T., Kaplan, I. R., Sakai, H. & Zak, I. *Chem. Geol.* **28**, 199-260 (1980).
- Knoll, A. H. *Origins of Life* **9**, 313-327 (1979).
- Fuchs, G., Thauer, R., Ziegler, H. & Stichler, W. *Arch. Mikrobiol.* **120**, 135-139 (1979).
- Deines, P. in *Handbook of Environmental Isotope Geochemistry* Vol. 1 (eds Fritz, P. & Fontes, J. C.) 329-406 (Elsevier, Amsterdam, 1980).

Evolution of radio structure in quasars: a new probe of protogalaxies?

P. D. Barthel* & G. K. Miley†‡

* Owens Valley Radio Observatory, California Institute of Technology 105-24, Pasadena, California 91125, USA

† Space Telescope Science Institute, Homewood Campus, Baltimore, Maryland 21218, USA

Analysis of new high-resolution radio observations of quasars with redshifts >1.5 shows that these powerful extragalactic radio sources have epoch-dependent physical properties. Distant quasars have a more bent, distorted appearance and are smaller than those nearby. The ambient medium, interacting with the radio jets, is postulated as playing an important role in producing these epoch-dependent effects. Optical absorption line studies are important in this respect. These observations provide an important new probe of young galaxies and their formation.

THE radio morphologies of quasars contain unique information about the physics of radio sources, the nature and the evolution of the environment in which they are located, and the geometry of the Universe. Since the mid-1960s, there have been several attempts to study redshift-dependence of radio sizes of quasars as a test of cosmological size evolution¹⁻³, showing indications of size evolution in the form of a deficit of large linear sizes at high redshift. The main problems in such studies have been to disentangle the various geometrical and physical effects from each other and from those of observational selection. Before the advent of the Very Large Array (VLA)⁴, the angular resolutions and sensitivities available permitted scant morphological information to be obtained for only a handful of high redshift ($z > 1.5$) quasars. More serious, because of the Malmquist bias, the few mapped high-redshift quasars were also systematically more radio-luminous than the studied low-redshift quasars. It was therefore impossible to separate unambiguously the effects of redshift from those of intrinsic luminosity. Here we present results of the first study with the VLA designed to overcome some of these problems. The VLA, with its combination of high sensitivity and sub-arcsecond resolution, can provide data on lower luminosity sources at high redshift, allowing an attempt to separate the effects of luminosity and redshift. In addition, the spatial resolution of the VLA is sufficient to allow more sophisticated morphological parameters than mere sizes to be determined at high redshift. It was hoped that analysis of such parameters would help to disentangle the various redshift-dependent physical and geometrical effects. Before we began this project, the extended radio structure of only a few high redshift quasars, mainly 3C sources, was known. To extend this number we mapped some hundred powerful quasistellar radio sources, with steep or previously unknown radio spectra, from the Cambridge 4C, Parkes (PKS), Bologna (B2), Molonglo (MC), and Westerbork (W) catalogues, at sub-arcsecond resolution. The quasars we observed had redshifts between 1.5 and 3.2 and monochromatic luminosities at 5 GHz (emitted frequency) between 10^{26} and 10^{28} W Hz⁻¹, adopting a Hubble constant $H_0 = 75$ km s⁻¹ Mpc⁻¹ and deceleration parameter $q_0 = 0.5$. The resulting 5 GHz radio maps are shown in Barthel *et al.*⁵ (BMSL hereafter); in this paper we draw attention to one of the most important results of our work, namely the epoch-dependence of the radio morphologies.

Observations, data analysis

The starting point for our work was the quasar compilation of Hewitt & Burbidge⁶. We extracted all quasars with redshift $z > 1.5$ (subsequently referred to as 'high redshift' quasars)

known to exhibit radio emission and having $\delta_{1950} > -30^\circ$. Where possible, radio spectral indices were obtained from the literature, and (because extended radio emission invariably connotes a steep radio spectrum) a sample of some hundred sources having either steep ($\alpha \geq 0.6$, where spectrum $S_\nu \propto \nu^{-\alpha}$) or unknown spectra was constructed. Most of the sources for which no spectral information was available were Molonglo (MC2, MC3 and MC5), Bologna (B2), and Westerbork (W) quasars, with only one measured spectral point (flux density at 408 or 1420 MHz). It should be stressed here that our spectral criterion, radio spectral index from 1.4 to 5 GHz, $\alpha_{1.4}^{5.0} \geq 0.6$, converts to $\alpha_s^{15} \geq 0.6$ for a typical high redshift quasar in its emitted frame: we have restricted ourselves to truly steep spectrum sources. For a few bright sources, high-resolution maps were already available in the literature. We observed the remaining quasars with the VLA (configuration A, 4885 MHz, 50 or 100 MHz bandwidth) in snapshot mode, during several observing sessions between 1982 and 1986. The typical dynamic range obtained in the maps was 200:1, and typical angular resolution was 0.5 arcsec, which corresponds to about 3 kpc linear resolution. Details of the data reduction procedures as well as the radio maps can be found in BMSL⁵. Total source flux densities at 5 GHz were derived from the data and used to obtain spectral indices for those quasars with previously unknown spectra. Several of these had insufficiently steep spectra to be included in our final steep spectrum, high redshift sample. For many quasars the position of the associated optical object was not known with sufficient accuracy for useful comparison with our radio data. Using Palomar Observatory Sky Survey plates we measured the positions of these objects to an accuracy of 0.3–0.6 arc sec (standard error). The new positions, given in BMSL, were used in subsequent analysis of the data. To compare these high-redshift data we have constructed a similar sample of lower redshift quasars, using existing data in the literature. In constructing this sample we used the following four criteria: (1) $z < 1.5$; (2) $\alpha_s^{15} \geq 0.6$ for $z < 0.7$, and $\alpha_{2.7}^{10.7} \geq 0.6$ for $0.7 < z < 1.5$, to comply with the high- z spectral index constraint, including the K-correction; (3) availability of maps and/or detailed structural information at few arc sec resolution (1.4, 2.7, 5 or 15 GHz); (4) luminosity $P_5 \geq 10^{26}$ W Hz⁻¹, as the lowest luminosity in the high-redshift sample is 10^{26} W Hz⁻¹. In summary, we will compare samples of quasars with $\alpha_s^{15} \geq 0.6$ and $P_5 \geq 10^{26.0}$ W Hz⁻¹, in the emitted frames.

Many morphological parameters were defined and investigated. We will concentrate on two of the most important parameters, namely observed angular and linear sizes and bending of the quasar radio morphologies. The Largest Angular Size (LAS) is defined³ as the angular distance between the source extremities, and is straightforward to measure. We obtain the Largest Linear Size (LIN) by multiplying the quasar LAS value

‡ Permanent address: Leiden Observatory, PO Box 9513, NL-2300 RA Leiden, The Netherlands.

Table 1 The high-redshift quasar sample

Quasar	Other name	z	S_5^*	Spectral index, $\alpha^\ddagger$	$\log P_5^\dagger$	Morphology	LAS (arc sec)	LIN (kpc)	Bending (degrees)	Ref.
0017+154	3C9	2.012	550	1.0	27.9	T	13	71	21	25
0032+423	4C42.01	1.588	120	1.0	27.0	C	5	27		5
0033+079	4C08.04	1.578	130	1.2	27.1	T	5	28	26	5
0038-019	4C-02.04	1.690	280	1.1	27.4	T	21	120	12	5
0044-055	4C-05.03	1.869	140	0.8	27.1	SSC	1	5		5
0046-067	4C-06.04	2.063	140	0.8	27.2	D1	9	46	30	5
0051+291	4C29.01	1.828	240	0.6	27.2	D2	4	20		5
0109+176	4C17.09	2.157	120	1.3	27.4	T	14	75	59	5
0225-014	4C-01.11	2.037	150	1.1	27.4	T	17	92	35	5
0238+100	MC5	1.816	70	1.1	26.9	T	18	100	3	5
0316-203	MC	2.880	40	0.6	26.8	SSC	0.2	1		5
0352+123	4C12.17	1.616	230	1.0	27.3	T	9	51	46	5
0404+177	4C17.22	1.712	180	1.1	27.3	D2	8	45		5
0424-131	PKS	2.165	330	0.9	27.7	SSC	0.1	0.5		5
0445+097	4C09.17	2.110	390	0.8	27.7	D2	3	18		5
0549-213	MC	2.245	190	0.8	27.4	D2	4	19		5
0553-205	MC	1.544	50	0.8	26.5	T	8	45	8	5
0730+257	4C25.21	2.686	230	0.8	27.6	T	10	50	32	5
0730+659	W1	1.937	30	0.8	26.5	T	20	110	5	5
0747+613	O1680	2.492	350	0.8	27.8	SSC	0.4	2		5
0751+298	4C29.27	2.106	150	0.8	27.2	C	1.5	8		5
0758+120	MC5	2.660	50	0.8	27.0	T	10	51	10	5
0802+103	3C191	1.956	490	1.1	27.8	T	5	27	1	5
0835+580	3C205	1.534	600	1.1	27.7	T	18	102	0	5
0836+195	4C19.31	1.691	160	0.7	27.0	T	33	185	13	50
0843+136	4C13.39	1.875	180	0.8	27.2	D2	2	11		5
0848+155	O1180	2.010	220	0.7	27.3	SSC	0.8	3		5
0856+124	MC5	1.760	45	0.8	26.6	SSC	1.8	9		5
0926+117	4C11.32	1.754	140	0.9	27.1	T	8	45	20	5
0941+261	OK270	2.910	350	0.8	27.9	D2	1.5	7		5
1023+067	3C243	1.699	220	1.2	27.4	T	15	79	6	5
1055+499	5C02.56	2.390	100	0.8	27.2	SSC	2	10		5
1153+317	4C31.38	1.557	950	0.9	27.8	SSC	1	6		5
1158+122	MC2	2.018	20	0.7	26.3	D2	2	11		5
1214+106	MC2	1.884	40	1.0	26.7	T	28	155	31	5
1218+339	3C270.1	1.519	1,000	0.8	27.8	T	10	54	33	51
1221+114	MC2	1.755	120	0.8	27.0	SSC	1.5	8		5
1226+105	MC2	2.296	190	1.0	27.5	T	6	32	58	5
1258+404	3C280.1	1.659	350	1.1	27.5	T	23	130	12	52
1308+182	4C18.36	1.689	110	1.0	27.0	T	11	62	10	5
1311-270	PKS	2.195	220	0.9	27.5	T	19	100	6	5
1318+113	4C11.45	2.171	800	0.8	28.0	T	7	38	20	5
1323+655	4C65.15	1.618	180	1.0	27.2	T	9	50	52	5
1334+119	MC2	1.760	120	0.8	27.0	D2	11	61		5
1345+584	4C58.27	2.039	200	0.8	27.3	T	15	82	69	5
1354+258	PKS	2.032	90	0.7	26.9	T	14	76	41	5
1456+092	MC	1.991	270	0.7	27.4	D2	2	11		5
1511+103	MC2	1.546	50	1.1	26.6	T	35	200	22	5
1540+180	4C18.43	1.662	250	0.6	27.1	T	7	36	53	5
1554-203	MC	1.945	60	0.8	26.8	T	23	126	7	5
1557-199	MC	1.590	20	0.8	26.1	SSC	1	6		5
1559+173	4C17.65	1.944	280	0.7	27.4	D2	3	16		5
1602-001	4C-00.63	1.625	380	0.8	27.4	T	25	141	10	5
1602+576	4C57.27	2.850	400	0.6	27.8	SSC	0.2	1		5
1606+289	4C28.40	1.989	160	1.2	27.4	D1	30	164	5	53
1629+120	4C12.59	1.782	740	0.7	27.7	D2	2	11		5
1629+680	4C68.18	2.475	370	0.7	27.7	SSC	0.2	1		5
1634+176	MC3	1.897	100	1.2	27.2	T	7	39	2	5
1658+575	4C57.29	2.173	110	1.5	27.5	T	6	29	8	5
1702+298	4C29.50	1.927	570	0.8	27.7	C	4	19		5
1732+160	4C16.49	1.880	250	1.2	27.5	T	17	94	4	5
1816+475	4C47.48	2.225	150	1.1	27.4	T	6	32	12	5
1857+566	4C56.28	1.595	230	1.1	27.3	T	30	170	12	5
2120+168	3C432	1.805	400	1.1	27.6	T	16	90	11	54
2146-133	PKS	1.800	490	0.9	27.7	T	4	19	12	5
2149+212	4C21.59	1.534	370	0.6	27.2	SSC	2	10		5
2150+053	4C05.81	1.979	310	1.1	27.6	T	17	93	0	5
2156+297	4C29.64	1.753	450	0.8	27.5	SSC	1	6		5
2158+101	4C10.67	1.730	200	0.8	27.2	SSC	1	6		5
2209+152	MC3	1.502	70	0.8	26.6	T	15	85	10	5
2222+051	4C05.84	2.323	240	0.9	27.6	T	3.5	18	25	5
2223+210	PKS	1.960	1,050	0.7	28.0	SSC	1	5		5
2248+192	4C19.74	1.806	250	1.1	27.5	T	7	36	7	5
2249+185	3C454.0	1.757	800	0.8	27.8	SSC	1	6		5
2251+244	4C24.61	2.328	870	0.6	28.0	SSC	0.2	1		5
2332+489	OZ453.7	1.534	130	1.3	27.1	T	35	198	24	5
2338+042	4C04.81	2.594	450	1.0	28.0	T	3	15	31	5
2345+061	4C06.76	1.546	330	0.6	27.3	C	1	6		5
2353+154	MC3	1.801	500	0.7	27.6	SSC	0.05	0.2		5
2354+144	4C14.85	1.810	390	1.0	27.6	T	11	61	6	5

* Measured flux density at 5 GHz (mJy). † K-corrected monochromatic luminosity at 5 GHz (W Hz^{-1}). ‡ Radio spectral index, from ~5 to 15 GHz emitted frequencies.

Table 2 The low-redshift quasar sample

Quasar	Other name	z	S_5^*	α^*	$\log P_5^*$	Morphology	LAS*	LIN*	Bending*	Ref.
0127+233	3C43	1.459	1,080	0.8	27.8	T	3	17	96	55
0133+207	3C47	0.425	1,100	1.0	26.7	T	70	309	5	56
0134+329	3C48	0.367	5,350	1.0	27.2	SSC	1	4		57
0137+012	4C01.04	0.260	830	0.6	26.0	T	35	105	1	19
0210+860	3C61.1	0.184	1,900	0.9	26.1	D1	190	504	6	58
0229+341	3C68.1	1.238	830	1.2	27.6	D1	58	333	4	58
0349-146	3C95	0.614	600	0.9	26.7	T	114	583	2	59
0350-073	3C94	0.962	790	1.3	27.4	D1	42	237	1	59
0429+415	3C119	0.408	3,420	0.7	27.1	SSC	0.04	0.2		57
0518+165	3C138	0.760	4,100	0.8	27.7	SSC	1	5		57
0538+498	3C147	0.545	8,100	1.0	27.8	SSC	1	5		57
0704+384	4C38.20	0.579	275	0.9	26.3	T	22	110	1	59
0710+118	3C175	0.768	630	1.0	27.0	T	50	272	6	60
0740+380	3C186	1.063	380	1.3	27.2	SSC	2	9		61
0758+143	3C190	1.195	810	0.9	27.5	C	4	23		55
0809+483	3C196	0.871	4,300	1.1	28.0	D1	6	31	10	62
0833+654	3C204	1.112	300	0.9	27.0	T	36	206	2	59
0838+133	3C207	0.684	1,400	0.6	27.1	T	12	63	21	<i>a</i>
0850+140	3C208	1.110	550	1.2	27.3	T	14	80	3	56
0855+143	3C212	1.048	880	0.6	27.3	T	10	57	0	<i>b</i>
0903+169	3C215	0.411	450	1.1	26.3	T	45	195	17	19
0922+149	4C14.31	0.896	175	0.9	26.6	T	40	224	11	59
0927+362	3C220.2	1.157	600	0.8	27.3	T	9	49	1	60
0937+391	4C39.27	0.618	215	0.9	26.3	T	50	256	4	59
0957+003	4C00.34	0.907	330	0.6	26.7	T	36	200	22	19
1100+772	3C249.1	0.311	750	0.8	26.2	T	23	86	24	63
1111+408	3C254	0.734	770	1.2	27.1	T	14	75	17	64
1132+303	3C261	0.614	370	0.9	26.5	D2	8	41		61
1137+660	3C263	0.652	1,000	1.3	27.1	T	45	234	1	65
1156+631	4C63.15	0.594	260	0.9	26.3	D1	58	293	8	66
1206+439	3C268.4	1.400	600	1.1	27.6	T	11	63	15	63
1210+133	4C13.46	1.137	760	0.7	27.3	SSC	0.5	3		67
1213+538	4C53.24	1.065	920	1.2	27.5	T	33	188	5	64
1215+643	4C64.15	1.288	300	0.7	27.1	D1	10	57	5	66
1241+166	3C275.1	0.557	1,070	0.7	26.9	T	17	84	25	51
1250+568	3C277.1	0.321	1,000	1.1	26.4	SSC	2	6		57
1328+254	3C287	1.055	3,250	0.8	27.9	SSC	0.04	0.2		57
1328+307	3C286	0.849	7,450	0.7	28.1	SSC	0.03	0.2		57
1340+606	3C288.1	0.961	400	1.3	27.1	D1	6	34	7	68
1416+067	3C298	1.436	1,450	0.7	27.8	SSC	1.5	8		55
1458+718	3C309.1	0.905	3,750	0.8	27.9	T	2	10	34	57
1502+602	3C311	1.022	470	1.0	27.1	SSC	0.5	3		21
1542+210	3C323.1	0.264	920	1.1	26.2	T	68	230	4	68
1606+180	4C18.47	0.346	1,500	0.7	26.6	T	15	60	4	19
1618+177	3C334	0.555	580	0.6	26.5	T	17	85	17	19
1622+238	3C336	0.927	800	1.2	27.3	D1	22	124	3	68
1634+268	3C343	0.988	1,480	1.3	27.7	SSC	0.04	0.2		57
1634+269	3C342	0.561	320	0.9	26.4	T	39	193	8	59
1704+608	3C351	0.371	1,200	1.0	26.6	T	58	239	9	59
1828+487	3C380	0.692	6,500	0.7	27.8	C	7	37		69
2044-027	3C422	0.942	1,000	0.9	27.4	SSC	1	6		19
2135-147	PKS	0.200	1,400	0.7	26.1	T	150	422	4	59
2252+129	3C455	0.543	900	1.3	26.9	SSC	3	15		61
2349+327	4C32.69	0.659	220	0.8	26.3	T	62	324	13	70

* As in Table 1.

a J. F. C. Wardle, personal communication.*b* R. A. Laing, personal communication.

by the corresponding angular size distance, using the quasar emission-line redshift and the adopted cosmology. The bending B is defined as the complement of the angle between the lines from the peaks in the source extremities to the radio core, or in its absence the optical object, and has been measured for D1 (double with no detectable core emission) or T (triple) sources. The bending is zero for a source in which core and lobes are colinear, and undefined for one-sided (D2) sources. The final high-redshift sample comprises 80 quasi-stellar radio sources with $z > 1.5$ and radio spectral index $\alpha_{1.4}^5 \geq 0.6$ (in some cases $\alpha_{0.408}^5$). The final low-redshift comparison sample consists of 54 steep spectrum quasi-stellar radio sources with $z < 1.5$. Tables 1 and 2 list observed and derived parameters of interest, for both samples. We have used the morphological classification SSC (steep spectrum core) for all the sources which were too small to be classified as double or triple, and the classification C for (six) complex sources.

The quasars in our steep spectrum, high-redshift sample have redshifts ranging from 1.502 (2209+152) to 2.910 (0941+261),

luminosities from $1.3 \times 10^{26} \text{ W Hz}^{-1}$ (1557-199) to $10^{28} \text{ W Hz}^{-1}$ (1318+113, 2223+210, 2251+244, and 2338+042) and linear sizes from $< 1 \text{ kpc}$ (SSCs) to 200 kpc (1511+103); the median values for redshift and luminosity are 1.878 and $10^{27.4} \text{ W Hz}^{-1}$, respectively. Bending $> 20^\circ$ appears frequently. The low-redshift quasars range from $z = 0.184$ (0210+860) to 1.459 (0127+233), from $P_5 = 10^{26} \text{ W Hz}^{-1}$ (0137+012) to 1.2×10^{28} (1328+307) W Hz^{-1} , and in linear size from $< 1 \text{ kpc}$ (SSCs) to $\sim 600 \text{ kpc}$ (0349-146); the median values for redshift and luminosity are 0.747 and $10^{27.1} \text{ W Hz}^{-1}$, implying somewhat different luminosity distributions for the two samples. In Fig. 1 we show the redshift-luminosity distributions for the combined samples, and these distributions are indicative of a luminosity-redshift correlation to some extent. Both samples, however, cover two decades in intrinsic luminosity.

A quick comparison of the data suggests that larger linear sizes occur at low redshift, and that bending may be less prominent compared to the high-redshift quasars. Also, the fraction of one-sided (D2) sources appears lower at low redshift. We

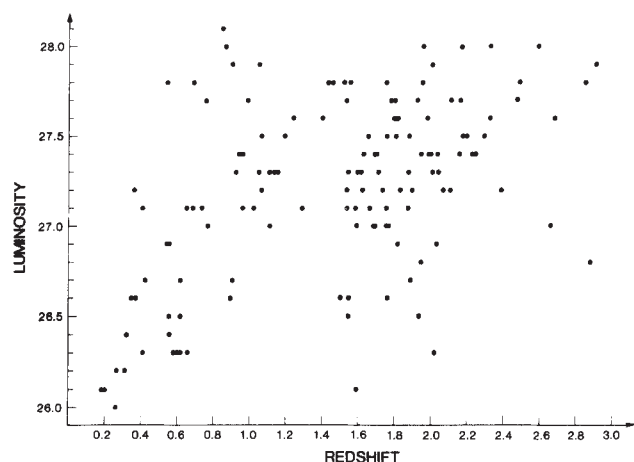


Fig. 1 Distribution in the redshift-luminosity plane for the combined sample of 134 steep spectrum quasars.

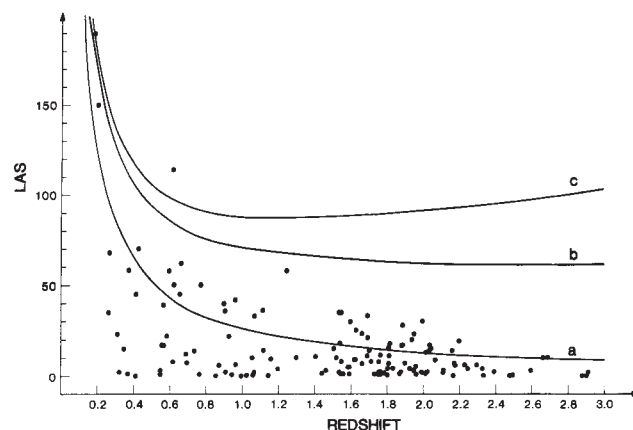


Fig. 2 Largest angular size (arc sec) vs. redshift for the combined sample of 134 steep spectrum quasars. The lines show the behaviour of a 500-kpc rigid rod in three cosmological models: (a) Euclidean universe; (b) Friedmann universe with $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 0.05$; (c) Friedmann universe with $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 0.5$.

will quantify these suggestions and investigate their origin below. As mentioned already, the main problem in the analysis is to separate the effects of luminosity and redshift, since these variables are correlated in flux density limited samples. The important difference with earlier work is that we can now attempt to make this separation, and ask whether the above mentioned trends are luminosity rather than redshift dependent.

Angular sizes. Figure 2 gives the LAS-redshift plot for our combined data set of 134 steep spectrum quasars. For comparison, we also show the behaviour of a 500-kpc rigid rod in various universe models. The appearance of a decreasing upper envelope in Fig. 2 agrees with a cosmological origin for the quasar redshifts; the deficit of large linear sizes at high redshift (see also the next section) in expanding universe models is evident. This evidence relies fairly heavily on the three low-redshift angular size points >100 arc sec, but an extensive radio galaxy data base at these low redshifts⁷ justifies our choice of a 500-kpc standard rod.

The remainder of this section is concerned with intrinsic parameters, and in the separation of redshift-dependence from luminosity-dependence. We assume a Hubble constant of $75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and a deceleration parameter of 0.5; smaller q_0 increases the inferred intrinsic luminosity and linear size values for high redshift objects. The sensitivity of our conclusions to the choice of q_0 , however, is small, and will be commented on below.

Linear sizes. To determine whether the apparent deficit of large linear sizes at high redshift (see Fig. 2) is an effect of redshift or intrinsic luminosity, we have divided the redshift-luminosity plane into nine sections. The redshift bins were 0.301–1.100, 1.101–1.900 and 1.901–2.700 (low, medium, high redshift) and the log P bins were 26.0–26.6, 26.7–27.3 and 27.4–28.0 (low, medium, high luminosity). This selection excludes eight sources from both samples, having $z < 0.301$, $z > 2.700$, or log $P > 28.0$. The remaining 126 quasars were distributed over the nine sections, and median values for redshift, luminosity, and linear size, for each section were calculated. The twelve one-sided (D2) sources in the samples were included in the analysis after first doubling their linear sizes. The linear size parameter used was therefore in all cases proportional to the distance from the nucleus to the source extremities, the physically relevant quantity. To investigate the relative importance of both redshift and intrinsic luminosity on the observed linear size, we then fitted equations of the form

$$\log(\text{LIN}) = b_0 + b_1 \log(1+z) + b_2 \log P$$

to the median value data in the nine sections, and tested the significance of the partial regression coefficients b_1 and b_2 . On

the basis of the resulting two-dimensional regression equations it appears that no significant statements can be made on the influence of either redshift or luminosity on quasar linear size. This is largely due to the (still) non-uniform coverage of the redshift-luminosity plane and small-number statistics which, given also the fact that we investigate projected morphologies, result in poorly determined median parameter values in some sections. Moreover, the linear size distributions of the sources in each section differ considerably, in the sense that sources of galactic and supergalactic size are not equally represented in each section (see also ref. 8). We are interested, however, in the maximum linear size that a source of given redshift and luminosity can attain. From the two-dimensional regression of LIN_{max} in the nine sections we find

$$\begin{aligned} \log(\text{LIN}_{\text{max}}/\text{kpc}) &= 3.81 - (1.46 \pm 1.36) \log(1+z) \\ &\quad - (0.03 \pm 0.28) \log P \end{aligned}$$

The errors in this equation are standard errors of the partial regression coefficients, the significance of which can be tested using Student's t -test. Using a one-tailed test we can reject the null hypothesis, that LIN_{max} does not decrease with redshift, at the 18% level. The 80% confidence limits for this redshift regression coefficient are 0.23 and 2.69. The quasar luminosity, on the other hand, probably has no effect on the maximum linear size of the radio morphology. But the significance of these effects is not very high and will be dealt with again below. We can improve on the significance of the above mentioned regressions by studying the correlation of linear size separately with each of redshift or luminosity alone, regardless of the other, by performing a one-dimensional regression. This ensures more uniform binning, and consequently more reliable median values. Since in our data luminosity and redshift are still correlated to some extent (see Fig. 1), we do of course expect to find common trends. We therefore wish to determine the relative importance of the effects of redshift and luminosity. Using redshift bins $\Delta z = 0.5$ and luminosity bins $\Delta \log P = 0.3$, we find

$$\log(\text{LIN}_{\text{med}}/\text{kpc}) = 2.41 - (1.97 \pm 0.24) \log(1+z)$$

$$\log(\text{LIN}_{\text{med}}/\text{kpc}) = 19.16 - (0.64 \pm 0.11) \log P$$

The regression coefficients in these two equations differ from zero at the 0.06% and 0.1% significance level, respectively, using a one-tailed test. We indeed find common trends, with the redshift regression being the most significant one. One-

dimensional regression on LIN_{max} results in very similar equations. It therefore appears that the linear size of a quasi-stellar radio source evolves with redshift, in agreement with the results from the two-dimensional regression analysis. The above regression techniques, however, indicate that we cannot rule out a possible inverse correlation between linear size and intrinsic source luminosity. Indeed, a physical cause for such an inverse correlation can be envisaged: if an energetically powerful source is constrained within galactic dimension, a larger fraction of the source energy will become available as luminosity, compared to an extended source where the energy is also being used for source expansion. Such an inverse correlation was previously⁹⁻¹² proposed to be solely responsible for the deficit of large linear sizes at high redshift. We have shown that this is not true, the prime difference from earlier investigations being that we have been able to separate the effects of redshift and luminosity to a considerable extent. On the basis of the available data, we conclude that the linear size of quasi-stellar radio sources of constant intrinsic luminosity most likely evolves as $(1+z)^{-k}$, where $k=1.5-2$. Data obtained by Wills¹³ suggested similar evolution. To determine the extent of a probable inverse correlation between intrinsic luminosity and linear size, still larger samples are required, concentrating on low-luminosity, high-redshift quasars. Size evolution for samples of radio galaxies of constant luminosity has been investigated by Eales¹⁴ and Kapahi⁷. The former finds $\text{LIN}(z) \propto (1+z)^{-(1.45 \pm 0.4)}$, the latter $\text{LIN}(z) \propto (1+z)^{-(2 \pm 0.5)}$, for $q_0=0.5$. Our value for quasars agrees well with these findings. A direct correlation between luminosity and linear size has been shown to exist for nearby, low-luminosity radio galaxies ($22.5 < \log P < 25.5$) (refs 15-17). Such a direct correlation can be easily explained from the physics of source expansion combined with an intergalactic density evolving with $(1+z)^3$. We seen no evidence at all for such a direct correlation in our quasar data. If, as far as their radio properties are concerned, quasars blend in with radio galaxies, we have to conclude that such a direct luminosity-size relation does not continue above a certain critical source luminosity (and probably even reverses into an inverse relation). Oort *et al.*^{16,17} have also proposed that the cosmological size evolution for these low-luminosity radio galaxies may be considerably steeper: for $q_0=0$ they find $\text{LIN}(z) \propto (1+z)^{-(3.3 \pm 0.5)}$, but at this stage we do not regard the difference between the exponent of the strong source and weak source evolution as significant enough to warrant postulating luminosity-dependent size evolution. The main difference between the cumulative linear size functions for our high- and low-redshift samples is a tail towards large linear sizes in the low-redshift sample. This is the very reason for the size evolution established above. Furthermore, from comparison with uniformly distributed linear sizes, we conclude that neither the low- nor high-redshift quasars have such uniform distributions: about 25% of all steep spectrum quasars appear to have linear sizes less than 10 kpc (SSCs).

We have also searched for possible relations between source linear size and spectral index. It appears that the average (high frequency) spectral index tends to increase with increasing linear size, from about 0.8 to about 1.1, in both low- and high-redshift sample. This behaviour reflects the dominance of aged emission in the largest sources. The low- and high-redshift sample differ, however, in the average spectral indices for the SSC sources, with values of 0.91 and 0.74 respectively. This effect requires confirmation from larger samples, but has the intriguing implication of a decreasing energy burst rate in SSCs with cosmic epoch. **Bending.** Proceeding with the two-dimensional regression, we find an increase of radio source bending with redshift:

$$\log(\text{bending/degrees}) = 3.82 + (1.82 \pm 1.46)\log(1+z) \\ - (0.13 \pm 0.41)\log P$$

where we have used median values. Using a one-tailed test we can reject the null hypothesis that bending does not increase

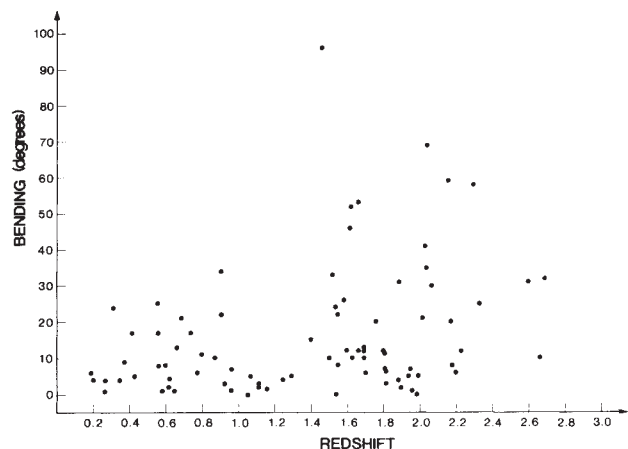


Fig. 3 Bending angle (degrees) as function of redshift for combined sample of 83 steep spectrum quasars.

with redshift at the 13% level. The 85% confidence limits for this redshift regression coefficient are 0.13 and 3.51. The quasar luminosity, on the other hand, definitively has a less significant effect on bending in the radio morphology. To improve on the statistics, we also performed one-dimensional regression, and found

$$\log(\text{bending/degrees}) = 0.43 + (1.73 \pm 0.43)\log(1+z)$$

$$\log(\text{bending/degrees}) = -6.95 + (0.29 \pm 0.11)\log P$$

The dependence of bending on redshift is significant at the 0.8% level, whereas luminosity is only significant at the 2.5% level. This very important result is shown again in Fig. 3, which shows the observed bending angle as a function of redshift for the combined samples. (An earlier version of this plot was given in ref. 18.) The most distorted quasars are 3C43(0127+233) at $z=1.453$, MC2 1226+105 at $z=2.296$, and 4C58.27 (1345+584) at $z=2.039$. Moreover, several complex and one-sided quasars in the high-redshift sample show pronounced bending in their morphologies, through angles up to 85° . Since we have defined the bending angle only for D1 and T sources, these quasars do not have bending entries in Table 2 and they do not appear in Fig. 3. We wish to stress again that we have been dealing here with truly steep spectrum quasars, having $\alpha_{15}^{15} \geq 0.6$ in their emitted frame. This is why some low-redshift quasars which were found to display considerable bending by Hintzen *et al.*¹⁹ do not appear in Fig. 3, since they have flatter spectra due to a dominant core. High dynamic range maps of similar high-redshift quasars are required to investigate possible evolution in the properties of flat spectrum quasars.

It is interesting to examine the high-redshift bending more closely. Twenty sources in the high-redshift sample have bending $\geq 20^\circ$. Of those 20, 7 show pronounced jet bending on the 5-10 kpc scale (distance from the radio core), 11 show non-collinearity between core and lobes or hotspots (at distances >25 kpc from the core), and 2 show bending within a radio lobe, at about 25 kpc from the core. In none of 12 cases where we can examine the bending in detail does there appear to be a morphological relationship between the two sides of the quasar; the straight side does not appear to know that the other side is bent.

Taken together, our data indicate epoch-dependent physical properties for extended quasi-stellar radio sources. As a final argument in favour of this proposal, we mention the fact that all the high-redshift sources in this study have intrinsic luminosities at 5 GHz (emitted frequency), P_5 , in excess of $10^{26} \text{ W Hz}^{-1}$, which for these steep-spectrum sources implies $P_{1.78} \geq 10^{27} \text{ W Hz}^{-1}$. *A priori*, apart from the SSC sources, we would therefore expect only edge-brightened double mor-

phologies, since this lower limit for the luminosity is an order of magnitude higher than that empirically derived by Fanaroff & Riley²⁰ for lower redshift edge-brightened double sources. Inspection of the morphologies in BMSL shows that this is definitely not the case.

Discussion

The apparent deficit of large quasars at high redshift can be attributed to an intergalactic or intracluster medium denser than occurs locally^{3,21}. On the simplest assumptions the average density will vary as $(1+z)^3$. Equating the internal pressure of a radio hotspot at the end of a plasma beam with the ram pressure in such an evolving IGM, and subsequently relating the hotspot advance velocity to the radio source linear size results in size evolution $\propto (1+z)^{-1.5}$. The dependence of density on redshift may be more severe for $z > 2$, since at early epoch host galaxies may well have denser, clumpier, and larger associated media. Evidence for the existence of such inhomogeneous high-redshift environments, both on intra- and intergalactic scale, will be discussed below. Quasar linear size scales may reflect the interface between a dense, gaseous galactic halo and a hot intergalactic medium, at the location of which the radio jet advance velocity will decrease²². Fall and Efstathiou²³ have argued that if galaxies form by dissipative collapse from a protogalactic halo, these halos should have been factors of ~ 4 – 8 times larger than their offspring galaxies. The size evolution might therefore be associated with evolving dense galactic halos. As parametrized above, the evolution is sensitive to the adopted q_0 -value. Referring to Fig. 2, linear size evolution $\propto (1+z)^{-k}$ with $k = 1$ – 1.5 , in a universe with $q_0 = 0.05$, would fit the data almost equally well. This also agrees with the radio galaxy evolution studies⁷ but as long as the linear size evolution as referred to above is not fully understood, conclusive statements on these matters would be very premature. Finally, the high fractional occurrence of subgalactic sized SSCs provides important evidence for the presence of host galaxies surrounding quasars.

We consider four possible mechanisms which might produce the observed increase in radio source bending at high redshift.

The observed bending angle is magnified compared with the intrinsic one due to the orientation of the radio source close to the line of sight²⁴. We do not expect this effect to be important, since our steep spectrum selection criterion should discriminate against possible Doppler boosting and small angles to the line of sight. Because, in a randomly oriented sample, only 3% of the members will be within 15° from the line of sight, we believe that certainly in our sample such effects are small (see also Lonsdale and Barthel²⁵).

The observed bending is due to gravitational lensing. Gravitational distortion of the images or radio sources by inhomogeneities in the universe²⁶ will be small, unless substantial but unseen, mass concentrations occur. Furthermore, if the lensing phenomenon were important, it would result in larger linear sizes at high redshift, which seems not to be the case. Finally, VLBI observations²⁷ have indicated that in several quasars the jet bends occur on very localized (10–100 pc) scales, which can be highly polarized²⁸. These findings argue against lens amplifications of intrinsically smooth bends, but we cannot rule out the effect and it clearly needs further investigation.

There are two possible effects which would cause physical bending of the radio jets. The first is changes in the positions and axes of the radio source production machine. It is possible that, as these quasars are close to the apparent redshift cut-off^{29,30} the axes of the nuclear jets are less stable. If galactic activity is triggered by interaction or merging with companion galaxies^{31–33}, repeated activity in a time of rich cluster formation could lead to the observed morphologies. Provided there is alternating ejection in these powerful quasars, such intrinsic mechanisms could explain the morphologies with non-collinear radio lobes.

The last and most plausible mechanism is interaction of the radio source with an epoch-dependent interstellar or intracluster medium (refs 18, 28, 34; Barthel *et al.*, in preparation). Strong evidence that interaction of the radio source with gas clouds can result in jet bending is provided by extranuclear line-emitting regions associated with extended radio sources in less distant radio galaxies^{35–37}. Van Breugel *et al.*^{36,37} estimate galactic ram pressures of $\sim 10^{-9}$ dyn cm⁻² in the low-redshift radio galaxies 3C293 and, particularly striking, Coma A, whereas in high-redshift quasars we calculate galactic ram pressures of $\sim 10^{-7}$ dyn cm⁻² to explain the observed radio-jet bending (ref. 28; Barthel *et al.*, in preparation). Such values indicate high galactic densities or high (turbulent) velocities. The fact that in many cases the bending does not appear to be correlated between different components of the same quasar argues strongly for these interaction phenomena. Some cases of observed bending can also be explained by buoyancy of the energy flows in the triaxial potential of large masses of hot gas, which one might expect in high redshift galaxies³⁸. As for bending by interaction of the radio source with the intracluster medium, we repeat that the average intracluster medium (ICM) density at high redshift will be about two orders of magnitude higher, but also more clumpy than locally. Massive, dense ICM clouds, such as have been postulated by Lonsdale and Barthel³⁹ may be able to deflect powerful radio jets. It is in this respect interesting to enquire how many pre-existing clouds would be needed to intercept a typical radio jet. Very crudely, if N_{cl} is the number of clouds per quasar, Ω_j the jet opening angle, and F_{bent} the observed fraction of bent quasars, then $N_{cl} \approx (\Omega_j/4\pi)^{-1} F_{bent}$. This implies a few hundred clouds per quasar. Since jet deflection arguments require deflecting masses of order 10^7 – $10^8 M_\odot$ (ref. 39), this implies a total mass in clouds of $\sim 10^{10} M_\odot$ per quasar. Such clouds, with sizes of a few kpc, would have column densities of $\sim 10^{21}$ cm⁻². Because this is comparable with the column densities observed from associated heavy element absorption line systems in QSOs (ref. 40) it is tempting to associate the clouds responsible for bending the radio source with associated absorption lines.

An extensive spectroscopic program currently being carried out by us using the Palomar 200-inch telescope supports this idea, and suggests typical distances for the metallic clouds of the order of 1–10 kpc (Barthel *et al.*, in preparation). We stress however that large samples are needed, because of differing covering factors. Both we, as well as Foltz *et al.*⁴¹ and Anderson *et al.*⁴², find in general a high fraction of strong associated absorption systems in the optical spectra of steep spectrum quasars at high redshift. A suggestive trend between radio compactness and the presence of associated absorption line systems⁴³ is also interesting in this respect.

Having pointed out the evidence for a dense, clumpy medium at high redshift, we consider its possible origin. The first possibility is that it is a relic of galaxy formation. Recent modelling of the physical conditions in the early universe^{44,45} postulates the presence of three types of gas condensation at redshifts 2–3, the epoch of our interest. These condensations include HII clouds, with masses 10^6 – $10^{8.5} M_\odot$, gas- and metal-rich dwarf galaxies with masses $10^{8.5}$ – $10^{9.5} M_\odot$, and elliptical and spiral galaxies, $\geq 10^{9.5} M_\odot$ in mass and with very extended dark haloes. The metal-rich dwarf galaxies become trapped in the potential wells of the large, massive (proto)galaxies and may trigger nuclear activity in the latter⁴⁶. These dwarf galaxies have been postulated to be responsible for the metallic line absorption systems commonly found in high redshift quasars. The very same dwarfs may be to a large extent responsible for the distorted appearance of high-redshift radio sources reported here. Furthermore, there is increasing evidence from numerical simulations^{47,48} that in general the medium in early epoch galaxies may be clumpy. Also, Rees (personal communication) has recently advocated a two-phase structure of extended protogalactic haloes, in which hot gas ($T \approx 10^6$ K and density $n \approx$

10^{-3} cm^{-3}) is in pressure balance with cooler gas condensations of density $0.1\text{--}1 \text{ cm}^{-3}$. An alternative for the origin of the clouds is that they are ejecta of the quasars, possibly following a broad absorption line (BAL⁴⁰) phase of the quasar in a radio quiet stage⁴⁹. Both in the dwarf galaxy and in the BAL ejecta scenarios, and possibly in the two-phase medium picture, we may expect common occurrence of distorted radio morphology and associated heavy element absorption, and provided this is the case, the radio maps will enable us to infer the spatial distribution of the absorbing clouds.

The properties of high-redshift radio sources are indicative of dense environments interacting with the radio jets. These

distant radio sources therefore provide a unique new probe of the young universe and the late stages of galaxy evolution. Combining our VLA, MERLIN and VLBI observations with spectroscopic and imaging data, we are in the process of exploiting this field extensively.

We acknowledge useful discussions with M. Fall, M. Rees and R. Weymann. This work was begun when P.D.B. was at the Leiden Observatory, The Netherlands. G.K.M. is affiliated to the Astrophysics Division, Space Science Department of the European Space Agency, Noordwijk, The Netherlands. The VLA is operated by Associated Universities, Inc., under contract with the National Science Foundation.

Received 16 November 1987; accepted 3 March 1988.

1. Miley, G. K. *Nature* **218**, 933-934 (1968).
2. Legg, T. H. *Nature* **226**, 65-67 (1970).
3. Miley, G. K. *Mon. Not. R. astr. Soc.* **152**, 477-490 (1971).
4. Thompson, A. R., Clark, B. G., Wade, C. M. & Napier, P. J. *Astrophys. J. Suppl.* **44**, 151-167 (1980).
5. Barthel, P. D., Miley, G. K., Schilizzi, R. T. & Lonsdale, C. J. *Astr. Astrophys. Suppl.* (in the press) (BMSL).
6. Hewitt, A. & Burbidge, G. *Astrophys. J. Suppl.* **43**, 57-158 (1980).
7. Kapahi, V. in *IAU Symp. 124: Observational Cosmology* (eds Hewitt, A., Burbidge, G. & Fang, L.) 251-266 (Reidel, Dordrecht, 1987).
8. Allington-Smith, J. R. *Mon. Not. R. astr. Soc.* **210**, 611-631 (1984).
9. Stannard, D. & Neal, D. S. *Mon. Not. R. astr. Soc.* **179**, 719-740 (1977).
10. Wardle, J. F. C. & Potash, R. I. *Am. N.Y. Acad. Sci.* **302**, 605-610 (1977).
11. Hooley, A., Longair, M. S. & Riley, J. M. *Mon. Not. R. astr. Soc.* **182**, 127-145 (1978).
12. Masson, C. R. *Astrophys. J.* **242**, 8-13 (1980).
13. Wills, D. *Astrophys. J. Suppl.* **39**, 291-315 (1979).
14. Eales, S. A. *Mon. Not. R. astr. Soc.* **217**, 179-203 (1985).
15. Gavazzi, G. & Perola, G. C. *Astr. Astrophys.* **66**, 407-416 (1978).
16. Oort, M. J. A., Katgert, P., Steeman, F. W. M. & Windhorst, R. A. *Astr. Astrophys.* **179**, 41-59 (1987).
17. Oort, M. J. A., Katgert, P. & Windhorst, R. A. *Nature* **328**, 500-501 (1987).
18. Miley, G. K. in *IAU Symp. 124: Observational Cosmology* (eds Hewitt, A., Burbidge, G. & Fang, L.) 267-271 (Reidel, Dordrecht, 1987).
19. Hintzen, P., Ulvestad, J. & Owen, F. *Astr. J.* **88**, 709-758 (1983).
20. Fanaroff, B. L. & Riley, J. M. *Mon. Not. R. astr. Soc.* **167**, 31P-35P (1974).
21. Wardle, J. F. C. & Miley, G. K. *Astr. Astrophys.* **30**, 305-315 (1974).
22. Gopal-Krishna & Wiita, P. J. *Mon. Not. R. astr. Soc.* **226**, 531-542 (1987).
23. Fall, S. M. & Efstathiou, G. *Mon. Not. R. astr. Soc.* **193**, 189-206 (1980).
24. Readhead, A. C. S., Cohen, M. H., Pearson, T. J. & Wilkinson, P. N. *Nature* **276**, 768-771 (1978).
25. Lonsdale, C. J. & Barthel, P. D. *Astr. J.* **94**, 1487-1492 (1987).
26. Blandford, R. D. & Jaroszynski, M. *Astrophys. J.* **246**, 1-12 (1981).
27. Barthel, P. D. *Astr. Astrophys.* **126**, 16-21 (1983).
28. Lonsdale, C. J. & Barthel, P. D. *Astrophys. J.* **303**, 617-623 (1986).
29. Osmer, P. S. in *IAU Symp. 119: Quasars* (eds Swarup, G. & Kapahi, V.) 447-453 (Reidel, Dordrecht, 1986).
30. Barthel, P. D. *loc. cit.* 181-184.
31. Heckman, T. M. *et al. Astrophys. J.* **311**, 526-547 (1986).
32. Yee, H. K. C. & Green, R. F. *Astrophys. J.* **319**, 28-43 (1987).
33. Djorgovski, S., Spinrad, H., Pedetty, J., Rudnick, L. & Stockton, A. *Astr. J.* **93**, 1307-1317 (1987).
34. Barthel, P. D. & Lonsdale, C. J. *Mon. Not. R. astr. Soc.* **205**, 395-402 (1983).
35. Heckman, T. M., Miley, G. K., Balick, B., Van Breugel, W. J. M. & Butcher, H. R. *Astrophys. J.* **262**, 529-553 (1982).
36. Van Breugel, W., Heckman, T., Butcher, H. & Miley, G. *Astrophys. J.* **277**, 82-91 (1984).
37. Van Breugel, W., Miley, G., Heckman, T., Butcher, H. & Bridle, A. *Astrophys. J.* **290**, 496-516 (1985).
38. Allan, P. M. *Astrophys. J.* **276**, L31-L33 (1984).
39. Lonsdale, C. J. & Barthel, P. D. *Astr. J.* **92**, 12-22 (1986).
40. Weymann, R. J., Carswell, R. F. & Smith, M. G. A. *Rev. Astr. Astrophys.* **19**, 41-76 (1981).
41. Foltz, C. B. *et al. Astrophys. J.* **307**, 504-534 (1986).
42. Anderson, S. F., Weymann, R. J., Foltz, C. B. & Chaffee, Jr, F. H. *Astr. J.* **94**, 278-288 (1987).
43. Swarup, G., Saikia, D. J., Beltrami, M., Sinha, R. P. & Salter, C. J. *Mon. Not. R. astr. Soc.* **220**, 1-18 (1986).
44. Ikeuchi, S. & Ostriker, J. *Astrophys. J.* **301**, 522-543 (1986).
45. Ikeuchi, S. & Norman, C. A. *Astrophys. J.* **312**, 485-496 (1987).
46. Silk, J. & Norman, C. *Astrophys. J.* **234**, 86-99 (1979).
47. Zurek, R. W., Quinn, P. & Salmon, J. K. *Astrophys. J.* (in the press).
48. Quinn, P. & Zurek, R. W. *Astrophys. J.* (in the press).
49. Briggs, F. H., Turnshek, D. A. & Wolfe, A. M. *Astrophys. J.* **287**, 549-554 (1984).
50. Saikia, D. J., Shastri, P., Sinha, R. P., Kapahi, V. K. & Swarup, G. *J. Astrophys. Astr.* **5**, 429-473 (1984).
51. Stocke, J. T., Burns, J. O. & Christiansen, W. A. *Astrophys. J.* **299**, 799-813 (1985).
52. Swarup, G., Sinha, R. P. & Saikia, D. J. *Mon. Not. R. astr. Soc.* **201**, 393-400 (1982).
53. Riley, J. M. & Pooley, G. G. *Mem. R. astr. Soc.* **80**, 105-137 (1975).
54. Feigelson, E. D., Isobe, T. & Kembhavi, A. *Astr. J.* **89**, 1464-1477 (1984).
55. Pearson, T. J., Perley, R. A. & Readhead, A. C. S. *Astr. J.* **90**, 738-755 (1985).
56. Burns, J. O., Basart, J. P., De Young, D. S. & Ghiglia, D. C. *Astrophys. J.* **283**, 515-528 (1984).
57. Van Breugel, W., Miley, G. & Heckman, T. *Astr. J.* **89**, 5-22 (1984).
58. Laing, R. A. *Mon. Not. R. astr. Soc.* **195**, 261-324 (1981).
59. Miley, G. K. & Hartsuiker, A. P. *Astr. Astrophys. Suppl.* **34**, 129-163 (1978).
60. Jenkins, C. R., Pooley, G. G. & Riley, J. M. *Mem. R. astr. Soc.* **84**, 61-99 (1977).
61. Potash, R. I. & Wardle, J. F. C. *Astr. J.* **84**, 707-717 (1979).
62. Lonsdale, C. J. & Morison, I. *Nature* **288**, 66-69 (1980).
63. Lonsdale, C. J. & Morison, I. *Mon. Not. R. astr. Soc.* **203**, 833-851 (1983).
64. Owen, F. N. & Puschell, J. J. *Astr. J.* **89**, 932-957 (1984).
65. Shone, D. L. *thesis* (Univ. Manchester, 1985).
66. Owen, F. N., Porcas, R. W. & Neff, S. G. *Astr. J.* **83**, 1009-1020 (1978).
67. Ulvestad, J., Johnston, K., Perley, R. & Fomalont, E. *Astr. J.* **86**, 1010-1035 (1981).
68. Pooley, G. G. & Henbest, S. N. *Mon. Not. R. astr. Soc.* **169**, 477-526 (1974).
69. Wilkinson, P. N., Booth, R. S., Cornwell, T. J. & Clark, R. R. *Nature* **308**, 619-621 (1984).
70. Potash, R. I. & Wardle, J. F. C. *Astrophys. J.* **239**, 42-49 (1980).

Geochemical integrations of the neutrino flux from stellar collapses

W. C. Haxton & Calvin W. Johnson

Institute for Nuclear Theory, Department of Physics, FM-15, University of Washington, Seattle, Washington 98195, USA

The galactic neutrino flux from stellar collapses may be measurable in geochemical experiments. Absorption cross-sections can be maximized and backgrounds from solar neutrinos and terrestrial antineutrinos minimized by exploiting neutrino capture reactions in which nuclei are excited above the threshold for particle breakup. Only a few practical candidate reactions exist; one of them, $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$, is being studied in a current solar neutrino experiment, and an interesting limit on the galactic stellar collapse rate may soon be obtained.

THERE have been several discussions of possible geochemical experiments to determine the long-term solar neutrino luminosity¹⁻⁴. In such experiments one measures the concentration of a long-lived isotope that has been produced in a deeply-buried ore body by neutrino interactions. The advantage over laboratory experiments is that much larger concentrations of daughter atoms can be produced over geological times; certain geochemi-

cal experiments can also yield information not obtainable from laboratory measurements, such as gauging the long-term thermal stability of the Sun in the $^{97}\text{Mo}/^{98}\text{Mo}$ experiment⁴. The disadvantage of this technique in general is the need to tolerate the backgrounds and chemical environments found in nature.

The recent supernova in the Large Magellanic Cloud, SN1987A, has generated considerable excitement. In standard

descriptions⁷ of supernovae, $\sim 99\%$ of the binding energy of the daughter neutron star is released in the form of neutrinos. The number and energy distribution of approximately 20 neutrino events recorded by the Kamioka and IMB water Cerenkov detectors^{5,6} were consistent with theoretical estimates of the $\bar{\nu}_e$ yield from a supernova. No detector was operating with enough sensitivity to determine the ν_e , ν_μ , $\bar{\nu}_\mu$, ν_τ or $\bar{\nu}_\tau$ fluxes.

The frequency of and energy released by collapsing stars are fundamental issues in stellar and galactic evolution, but present some special problems for observers. One may have to wait some time for a second event like SN1987A, and a great deal longer before sufficient statistics are gathered to characterize a 'typical' stellar collapse. From this perspective the possibility of a long-term geochemical integration of the galactic neutrino flux is intriguing. In this article we propose that a few nuclear reactions might be exploited for this purpose, and that one, $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$, may soon⁸ provide an interesting constraint on the frequency of stellar collapses within our galaxy.

Galactic neutrinos

According to the standard model⁷, an energy of $\sim 3 \times 10^{53}$ erg is carried off by neutrinos from a type II supernova. About 10% of these are ν_e s produced by the initial neutronization of the core; the remaining 90% are neutrinos and antineutrinos of all flavours, emitted in thermal distributions as the newly-formed neutron star cools, with the total energy per flavour approximately independent of the flavour. Because of their charged-current reactions, the average energies of the ν_e s and $\bar{\nu}_e$ s are lower than those of the muon and tauon neutrinos. The predictions of Wilson *et al.*⁹ are shown in Table 1 for the collapse of a $25M_\odot$ star. The (presumably) $\bar{\nu}_e$ events recorded by the Kamioka and IMB detectors^{5,6} are in reasonable accord with theoretical expectations: most fits to these data gave average temperatures $T = 4\text{--}5$ MeV and a total $\bar{\nu}_e$ energy of $\sim (0.4\text{--}0.8) \times 10^{53}$ erg (refs 10–13; R. W. Mayle and J. R. Wilson, preprint).

Table 1 Neutrino numbers and temperatures for the collapse of a $25M_\odot$ mass star with a $1.6M_\odot$ iron core⁸. T is the average temperature of the neutrinos

ν type	N_ν (10^{57})	T (MeV)	E_ν (10^{53} erg)
ν_e	4.5	4.7	1.1
$\bar{\nu}_e$	3.1	5.0	0.8
ν_μ	2.3	10.0	1.2
$\bar{\nu}_\mu$	2.3	10.0	1.2
ν_τ	2.3	10.0	1.2
$\bar{\nu}_\tau$	2.3	10.0	1.2

The rate of neutrino-producing stellar collapses is highly uncertain. A lower bound is the rate of type II supernovae in the galaxy, $\sim 0.02 \text{ yr}^{-1}$ (ref. 14). However, it is unclear what fraction of stellar collapses produces optically brilliant supernovae, and what fraction of these occurs in unobscured parts of the galaxy. Perhaps the most reasonable estimate of the collapse rate comes from the stellar death rate for massive stars. Bahcall and Piran¹⁵ have performed such calculations using a detailed model of the stellar content of the galaxy and lifetimes determined by conventional stellar evolution. For stars more massive than $10M_\odot$ (above which stars should eventually collapse rather than produce a white dwarf or type I supernova) they determined a rate of 0.09 yr^{-1} . We adopt this value in our calculations.

The reactions we discuss below will be excited by galactic neutrinos with energies ≥ 10 MeV. In addition to stellar collapses, there are two other known sources that could, in principle, contribute to the galactic neutrino flux. The first, the ordinary stellar neutrino flux integrated over the galaxy, is several orders of magnitude smaller than the solar flux. The second is the production of neutrinos by matter accreted onto

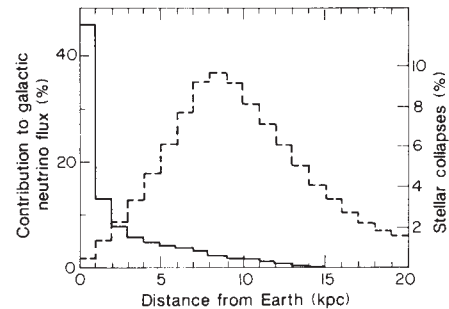


Fig. 1 Histogram of the number of stellar collapses (dashed line, right label) and the corresponding contributions to the galactic neutrino flux at Earth (solid line, left label) are given as a function of distance from Earth.

neutron stars and black holes. But these collapsed objects have already radiated away their binding energy as neutrinos. Thus, to produce a flux rivaling that from a stellar collapse, the accretion must at least double the mass of the original collapsed core. No astrophysical scenario has been proposed in which this will happen, or in which the accretion would yield many energetic (≥ 10 MeV) neutrinos. Estimates of the neutrino flux from black holes that may exist at the galactic centre give a flux about 6 orders of magnitude below that from stellar collapses¹⁶. Unless there is some unexpected astrophysical source, neutrinos from stellar collapse should dominate the galactic neutrino luminosity.

Type II supernovae seem to occur in the disc portion of galaxies. In our neutrino flux calculations we took the galactic density from the parameterization of Bahcall and Soneira¹⁷ as $\rho(r, z) \propto \exp(-|z|/z_0 + r/h)$ where r and z are cylindrical coordinates and the coordinate origin is the galactic centre. In this expression, $z_0 = 100$ pc and $h = 3.5$ kpc, and we place the Sun in the galactic plane 8 kpc from the centre. Varying the parameters z_0 and h within accepted limits ($70 \text{ pc} \leq z_0 \leq 150 \text{ pc}$, $3 \text{ kpc} \leq h \leq 4 \text{ kpc}$) changed the average flux at the Earth by no more than 5%. The average flux varies with the distance of the Sun from the galactic center by 13% per 0.5 kpc. The average total flux is $N_{58} 1.3 \times 10^4 (R/0.1 \text{ yr}^{-1}) \text{ cm}^{-2} \text{ s}^{-1}$, where R is the galactic stellar collapse rate and N_{58} is the mean number of neutrinos emitted per collapse in units of 10^{58} . One obtains the same flux if all galactic collapses are assumed to occur at an effective distance of 4.6 kpc from the Earth. Figure 1 shows the distribution of stellar collapses and their time-averaged contribution to the neutrino flux at Earth as a function of the distance from Earth.

In a geochemical experiment, the integration over time is finite. Do such experiments provide a reliable measure of the time-averaged galactic neutrino luminosity, or do the results depend sensitively on the chance occurrence of a collapse very near the Earth? To address this question we performed a number of Monte Carlo simulations, each involving 100,000 collapses distributed throughout the galaxy according to the density profile given above. (For a galactic stellar collapse rate of 0.09 yr^{-1} , each sample corresponds to a geochemical integration period of 1.1 Myr.) Collapses occurring beyond 100 pc generated 92% of the flux seen by the Earth, with very little variation (3%). Only a few (typically ~ 4) of the 100,000 events in each sample occurred closer than 100 pc, with fluctuations in the distribution of these events causing 1σ uncertainties in the total flux of about 10%. Thus a geochemical integration over 100,000 stellar collapses could determine the mean galactic neutrino luminosity to about 13%. The nearest collapse for a 100,000-event sample occurs at 71 ± 25 pc; a collapse positioned at the central value contributes only 4% to the total flux. Any estimate of the flux at Earth based on the likelihood of a single, nearby event, a procedure often followed in the literature, yields a result

that is too small by a factor of 25. We also note that an elevated flux due to a recent (and improbable) supernova within 30 pc can be ruled out by the absence of evidence for a mass extinction that should accompany the intense cosmic ray flux from such an event¹⁸.

Nuclear reactions of supernova neutrinos

There have been a number of proposals to exploit charged-current ν_e nuclear reactions in geochemical solar neutrino flux measurements, including $^{81}\text{Br}(\nu_e, e^-)^{81}\text{Kr}$ ($\tau_{1/2} = 2.1 \times 10^5$ yr) (ref. 1), $^{205}\text{Tl}(\nu_e, e^-)^{205}\text{Pb}$ ($\tau_{1/2} = 1.4 \times 10^7$ yr) (ref. 2), and $^{98}\text{Mo}(\nu_e, e^-)^{98}\text{Tc}$ ($\tau_{1/2} = 4.2 \times 10^6$ yr) (ref. 4). It is easy to demonstrate that a reaction of this type, in which a long-lived ($\tau_{1/2} \geq 10^5$ yr) daughter isotope is produced directly by the (ν_e, e^-) reaction, will be insensitive to galactic neutrinos because of the larger signal from solar neutrinos. The last reaction, $^{98}\text{Mo}(\nu_e, e^-)^{98}\text{Tc}$, is a good case to study because the high threshold eliminates contributions from pp, ^7Be , and CNO-cycle solar neutrinos, leaving the high-energy ^8B solar neutrinos (end point = 14 MeV) as the most copious solar source competing with galactic neutrinos. The cross-section averaged over the ^8B spectrum is approximately 3.0×10^{-42} cm² (the Gamow-Teller strength distribution for exciting bound states in ^{98}Tc was taken from forward-angle (p, n) calibrations¹⁹). The ^{37}Cl experiment²⁰ has established that the ^8B flux is weaker than predicted by the standard solar model²¹. For example, a low-metallicity non-standard solar model, adjusted to fit the ^{37}Cl counting rate, yields a flux $\phi(^8\text{B}) = 1.1 \times 10^6$ cm⁻² s⁻¹ (J. N. Bahcall, personal communication). The corresponding solar neutrino production rate of ^{98}Tc is 3.3 SNU (1 SNU = 10^{-36} captures per target atom per second). The corresponding cross-section averaged over a Fermi-Dirac neutrino distribution with a temperature $T = 5$ MeV is 53×10^{-42} cm². (Throughout this paper all neutrino cross-sections are evaluated in the allowed approximation.) Using a stellar collapse rate of 0.09 yr⁻¹ and an energy in ν_e s of 0.8×10^{53} erg, one obtains a galactic neutrino production rate of ^{98}Tc of 0.19 SNU, almost a factor of 20 smaller than the solar neutrino rate.

However, the ^{98}Mo case suggests an intriguing way to circumvent the solar neutrino background and, at the same time, benefit from a somewhat larger galactic neutrino cross-section. The idea is illustrated in Fig. 2a. The threshold for neutron emission by ^{98}Tc is 7.28 MeV. The excitation of states above this threshold in low-energy neutrino reactions will produce, with nearly 100% probability, the nucleus ^{97}Tc ($\tau_{1/2} = 2.6 \times 10^6$ yr). The analogue state ($E_x = 9.7$ MeV) and most of the Gamow-Teller strength lies in the continuum, so that, for $T \geq 2.6$ MeV, excitations of these states dominate the nuclear response to galactic neutrinos. The cross-section at $T = 5$ MeV is 150×10^{-42} cm², three times the discrete-state cross-section. In contrast, the ^8B solar neutrino continuum cross-section is suppressed by the limited phase space, giving $\sigma(^8\text{B}) = 0.29 \times 10^{-42}$ cm². Thus, repeating our

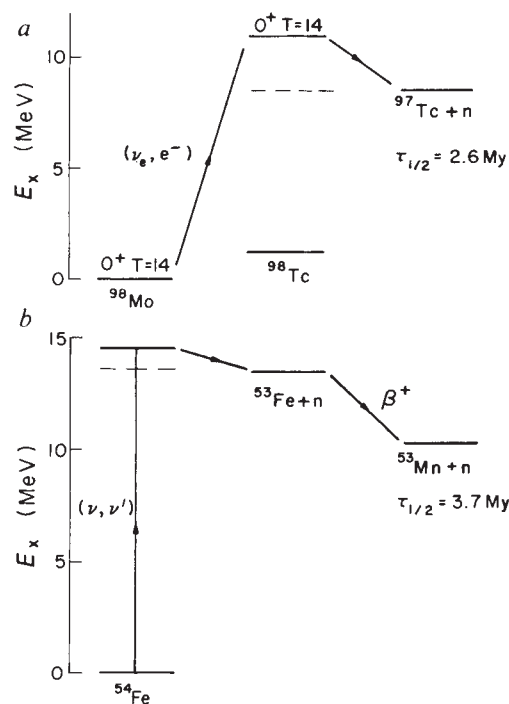


Fig. 2 Schemes for *a*, charged-current and *b*, neutral-current continuum nuclear reactions producing long-lived daughter nuclei. The dashed lines indicate thresholds for neutron emission. The position of the analogue state is shown in *a*.

earlier calculation, the galactic neutrino rate for producing ^{97}Tc via $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$ is 0.54×10^{-36} s⁻¹ per ^{98}Mo atom, 1.7 times the low-metallicity-model solar neutrino rate.

Thus it appears that the ^8B solar neutrino flux is not necessarily an overwhelming background to geochemical measurements of the galactic neutrino flux. But a second solar source is also of concern: despite the low flux ($\sim 1.6 \times 10^4$ cm⁻² s⁻¹, see ref. 21 for a discussion of the large uncertainty in this number), the high-energy neutrinos (18.8 MeV endpoint) generated in the β -decay of $^3\text{He} + p$ (hep neutrinos) readily excite high-threshold nuclear reactions. The hep neutrino cross-section for $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$ is 6.3×10^{-42} cm², so that the production rate is 0.10 SNU, 19% of the galactic rate given our assumed stellar collapse frequency and mean ν_e temperature. Note that the ^8B backgrounds can be made larger or smaller by adopting, rather than our low-metallicity 'consistent' solar model, another explanation of the ^{37}Cl puzzle. For instance, if matter-enhanced neutrino oscillations^{23,23} account for Davis's results²⁰, the 'adiabatic boundary' solution would yield a larger ^8B neutrino $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$ capture rate. The 'level-crossing boundary' solution, however, preferentially suppresses high-energy neutrinos so that the ^8B (and hep) continuum capture rates for ^{98}Mo would be much smaller.

We examined the table of isotopes to determine which high-threshold (ν_e, e^-) reactions are candidates for geochemical experiments. We required the parent nucleus to be reasonably abundant ($\geq 5\%$), the daughter to be long-lived ($\geq 10^5$ yr), and excluded obviously unsuitable reactions (such as noble gas parent nuclei). Table 2 suggests that very few opportunities exist. Only two reactions appear to deserve any consideration: $^{203}\text{Tl}(\nu_e, e^-)^{202}\text{Pb} + n$ and the ^{98}Mo reaction discussed above. From studies of ^{205}Tl as a pp-neutrino detector², one of its shortcomings can be anticipated: no deeply-buried, lead-poor thallium deposit is known. Also, the accumulation time for ^{202}Pb is somewhat short (~ 0.3 Myr). The ^{98}Mo reaction is also flawed: as ^{97}Mo is stable, solar neutrinos can produce ^{97}Tc by their interactions with this isotope. Nevertheless, because ^{97}Mo is less

Table 2 Charged-current ν_e reactions involving abundant parents ($\geq 5\%$), long-lived daughters ($\tau_{1/2} \geq 10^5$ yr), and high thresholds

Reaction	Parent abundance (%)	Daughter half-life (10^6 yr)	Threshold (MeV)	Comments
$^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$	24.1	2.6	8.96	$^{97}\text{Mo}(\nu_e, e^-)^{97}\text{Tc}$ by solar neutrinos troublesome
$^{100}\text{Mo}(\nu_e, e^-)^{99}\text{Tc} + n$	9.6	0.21	6.93	^{238}U spontaneous fission production of ^{99}Tc fatal
$^{130}\text{Te}(\nu_e, e^-)^{129}\text{I} + n$	34.5	16.0	6.91	Fission production via n (thermal) + ^{235}U appears fatal
$^{151}\text{Eu}(\nu_e, e^-)^{150}\text{Gd} + n$	47.9	1.8	6.96	Rare parent
$^{203}\text{Tl}(\nu_e, e^-)^{202}\text{Pb} + n$	29.5	~ 0.3	7.90	Daughter half life is short

Table 3 Charged-current ν_e reactions involving abundant parents ($\geq 5\%$), long-lived daughters ($\tau_{1/2} \geq 10^5$ yr), and high thresholds

Reaction	Parent abundance (%)	Daughter half-life (10^6 yr)	Threshold (MeV)	Comments
$^{11}\text{B}(\bar{\nu}_e, e^+)^{10}\text{Be} + n$	80.2	1.6	13.03	Thermal n capture on ^{10}Be fatal
$^{100}\text{Ru}(\bar{\nu}_e, e^+)^{99}\text{Tc} + n$	12.6	0.2	10.99	^{238}U spontaneous fission production of ^{99}Tc fatal
$^{136}\text{Ba}(\bar{\nu}_e, e^+)^{135}\text{Cs} + n$	7.5	3.0	10.33	$^{135}\text{Ba}(n, p)^{135}\text{Cs}$, $^{135}\text{Ba}(\bar{\nu}_e, e^+)^{135}\text{Cs}$ by absorption of terrestrial $\bar{\nu}_e$ s are very troublesome
$^{54}\text{Fe}(\bar{\nu}_e, e^+)^{53}\text{Mn} + n$	5.8	3.7	10.66	Also produced by neutral current (ν, ν')
$^{99}\text{Ru}(\bar{\nu}_e, e^+)^{98}\text{Tc} + n$	12.7	4.2	10.28	Rare parent

abundant (9.55%) than ^{98}Mo (24.13%) and because the feasibility of Tc extraction from molybdenum ores has been demonstrated⁸, this reaction is still of interest. We discuss the galactic neutrino 'background' to the current $^{97}\text{Mo}/^{98}\text{Mo}$ experiment in the next section.

The $(\bar{\nu}_e, e^+)$ reaction on $N > Z$ nuclei is suppressed relative to the (ν_e, e^-) reaction because of the nuclear Coulomb field, the absence of a Fermi transition, and the Pauli blocking of Gamow-Teller transitions. For medium-mass nuclei ($Z \approx 50$) the Coulomb effect is approximately an order of magnitude. The $(\bar{\nu}_e, e^+)$ reaction has the advantage however that anti-neutrino backgrounds are less severe. The principal background source, the Earth's radioactivity, produces a flux at the Earth's surface of $\sim 10^7 \text{ cm}^2 \text{ s}^{-1}$, with a spectrum extending from zero to $\sim 3.3 \text{ MeV}^{24}$. This background will not affect high-threshold reactions analogous to that of Fig. 2a. As noted earlier, such continuum reactions are generally favoured by their larger cross-sections for absorbing energetic neutrinos. The terrestrial anti-neutrino background will contribute to charged-current reactions to bound states, and its magnitude will depend on details of the Gamow-Teller distribution in the daughter nucleus. For nuclei near $Z = 50$, a low-lying transition with $\log ft \approx 5$ will lead to an antineutrino capture rate of 10^{-37} s^{-1} , a value comparable to the galactic antineutrino absorption rate for the continuum process of Fig. 2a. (The dominant absorption mechanism is antineutrino-induced resonant electron capture²⁴.) Thus, unless the low-lying Gamow-Teller strength is weak, low-threshold reactions may not be useful for detecting the galactic antineutrino flux.

The search for suitable $(\bar{\nu}_e, e^+)$ reactions is summarized in Table 3. In addition to processes like Fig. 2a, we also looked for daughters produced after proton emission: unlike the case with (ν_e, e^-) reactions, the product has Z different from the parent, so that chemical separation may be possible. However, this channel will only be appreciable for light nuclei or in cases where the proton emission threshold is significantly lower than the neutron threshold. Again the practical possibilities appear very limited. The only promising reaction is $^{54}\text{Fe}(\bar{\nu}_e, e^+)^{53}\text{Mn} + n$. We will see shortly, however, that supernova neutrinos produce ^{53}Mn principally by the neutral current reactions of muon and tauon neutrinos, rather than by the $(\bar{\nu}_e, e^+)$ reaction.

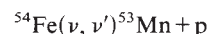
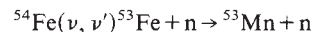
Neutral current reactions (ν, ν') are sensitive to neutrinos of all flavours, so the contributing galactic flux is much larger than for charged-current reactions. The higher average temperature of the muon and tauon neutrinos would in addition lead to substantially larger nuclear cross-sections for these species. Unfortunately, finding a geochemical 'signal' for such reactions requires some creativity. Transitions to bound states provide no signal, while those to continuum states will, after neutron emission, yield a nucleus of the same charge as the parent, thus making chemical separation of the long-lived daughter enormously difficult. Two other possibilities are more promising. In

Table 4 Neutral-current reactions involving abundant parents ($\geq 5\%$) and long-lived daughters ($\tau_{1/2} \geq 10^5$ yr)

Reaction	Parent abundance (%)	Daughter half-life (10^6 yr)	Threshold (MeV)	Comments
$^{11}\text{B}(\nu, \nu')^{10}\text{Be} + p$	80.2	1.6	11.23	Thermal n capture on ^{10}Be fatal
$^{54}\text{Fe}(\nu, \nu')^{53}\text{Fe} + p$	5.8	3.7	8.05	A possibility
$^{54}\text{Fe}(\nu, \nu')^{53}\text{Fe} + n$ $\xrightarrow{\beta^-} ^{53}\text{Mn} + n$			13.38	
$^{130}\text{Te}(\nu, \nu')^{129}\text{Te} + n$ $\xrightarrow{\beta^-} ^{129}\text{I} + n$	34.5	16.0	8.41	Fission production via n (thermal) + ^{235}U appears fatal
$^{151}\text{Eu}(\nu, \nu')^{150}\text{Eu} + n$ $\xrightarrow{\beta^- (89\%)} ^{150}\text{Gd} + n$	47.9	1.8	7.97	Rare parent

certain nuclei (with $Z \leq 20$, or with a proton threshold significantly lower than the neutron threshold) excitations into the continuum can be followed by appreciable proton emission, leading to a final nucleus of a different charge. Alternatively, as illustrated in Fig. 2b, the daughter nucleus formed following neutron emission could β -decay (or α -decay) to form a long-lived isotope of a different charge.

Table 4 summarizes the results of our search for a suitable reaction. Naively the most favourable appears to be the neutral current production of ^{53}Mn ($\tau_{1/2} = 3.7 \times 10^6$ yr) from ^{54}Fe (abundance 5.8%). There are two routes for this reaction:



The second will be important as the threshold for proton emission (8.85 MeV) is 4.5 MeV below the neutron emission threshold. To get a crude estimate of the cross-section we have performed a simple nuclear shell model calculation in which the states of ^{54}Fe are described as 2h or 1p3h excitations of a ^{56}Ni core. The Kuo-Brown interaction²⁵ was diagonalized in the 1f2p shell. The predicted energies of low-lying states in ^{54}Fe correspond quite well to experiment. Using the fluxes and temperatures of Table 1, we find a total neutral current capture rate of 1.25 SNU. In the shell model calculation, 61% of the strength is carried by states above the particle emission threshold, and 91% of the capture rate is due to muon and tauon neutrinos.

From a theoretical perspective this reaction seems ideal. The capture rate is large, comparable to the low-metallicity-model ^{37}Cl rate from ^8B solar neutrinos, and is dominated by heavy-flavour galactic neutrinos that have not yet been seen. The background due to solar neutrinos is very small. Although ^{54}Fe has an abundance somewhat less than ideal, iron ore is common and can be obtained in great quantities. Other practical issues are less clear: one must demonstrate that the large-scale chemical separation of Mn from iron ore is feasible, and that trace quantities of ^{53}Mn can be extracted from natural ^{55}Mn . Although there appear to be no obvious fatal backgrounds from natural radioactivity, this must be verified by careful calculations. The background from cosmic ray muons will be unusually severe: the existence of a neutral current channel for producing ^{53}Mn also results in large yields from muon scattering off ^{54}Fe (and ^{56}Fe), followed by neutron spallation. A measurement of the ^{53}Mn yield at the Earth's surface could be folded with known cosmic ray muon attenuation curves to determine the necessary overburden. It is very likely that the resulting depth requirement would rule out commercial iron ore deposits.

The $^{98}\text{Mo}/^{97}\text{Mo}$ experiment

Although the capture of galactic neutrinos by $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$ must compete with the solar neutrino reaction $^{97}\text{Mo}(\nu_e, e^-)^{97}\text{Tc}$, ^{97}Tc is of particular interest because it is the subject of an existing experiment. The motivation for the Los

Alamos $^{98}\text{Mo}/^{97}\text{Mo}$ solar neutrino experiment is to test the long-term thermal stability of the Sun by measuring the accumulation of ^{97}Tc and ^{98}Tc in ore from the deep Henderson Mine in Colorado^{4,8}.

A 10 kiloton sample of Henderson ore contains about 49 metric tonnes of MoS_2 and, assuming the low-metallicity-model 'consistent' fluxes, about 0.80×10^7 atoms of ^{97}Tc and 2.74×10^7 atoms of ^{98}Tc from ^8B neutrino reactions. The task of chemically isolating these trace quantities of Tc is simplified by the commercial processing of the ore. Flotation separation at the mine produces a concentrate that is 90 to 99% molybdenite, which is then shipped to the AMAX roasting plant at Ft. Madison, Iowa, for conversion to MoO_3 . The two roasters have a capacity of 50 tons per day, the quantity of concentrate needed for an experimental run. In the roasting process elements such as rhenium and technetium form volatile oxides that pass into the gas stream with high efficiencies. The gas stream is treated to remove dust, then cooled, scrubbed, and converted to sulphuric acid. The scrubbing removes rhenium, technetium, fluorides, chlorides, and selenium.

The initial experimental task is the quantitative recovery of technetium (and the chemically analogous rhenium) from the scrub stream. Resin ion-exchange columns for removing technetium from the scrub stream were first installed in October 1985, and six 12-hour runs were made in February 1986 when the entire stream was passed through four 12-inch columns operating in parallel. Based on rhenium recovery rates, better than 90% of the technetium was removed⁸.

Subsequent work has concentrated on extraction of Tc from the resin to produce an essentially massless specimen suitable for mass spectrometry. After molybdenum is eluted from the column, the resin is ashed. About 95% of the rhenium is volatilized during this stage. Further processing and cleanup of the 30 g sample results in a few-microlitre sample suitable for mass spectrometry. Both positive-ion and negative-ion mass spectrometry methods have been developed. The detection limits for these techniques now stand at 3×10^7 atoms and 5×10^6 atoms respectively for pure technetium samples⁸. It thus appears that the feasibility of the separate steps necessary for a successful experiment has been demonstrated.

This experiment seems to offer some near-term prospect of placing a constraint on the galactic neutrino flux. The key quantity is the ratio of the Tc produced by galactic neutrinos to that produced by solar neutrinos. In the case of ^{98}Tc , we showed earlier that this ratio is about 6% for the 'consistent' low-metallicity solar model, a supernova frequency of 0.09 yr^{-1} , and a ν_e -temperature of 5 MeV. The solar neutrino production of ^{97}Tc by exciting ^{98}Mo above the neutron emission threshold is 0.80 SNU. (Here and below the SNUs are normalized to the number of ^{97}Mo , not ^{98}Mo , atoms.) The solar neutrino rate for $^{97}\text{Mo}(\nu_e, e^-)^{97}\text{Tc}$ is not known because the Gamow-Teller distribution in ^{97}Tc has not yet been calibrated in (p, n) reactions. In the absence of a measurement, it is not unreasonable to scale the ^{98}Mo cross-section by the appropriate ratio of ($N-Z$) factors, in accordance with the naive sum rule. This yields a capture rate of 3.1 SNU, and thus a total capture rate of 3.9 SNU. The galactic neutrino production rates for $^{98}\text{Mo}(\nu_e, e^-)^{97}\text{Tc} + n$ and

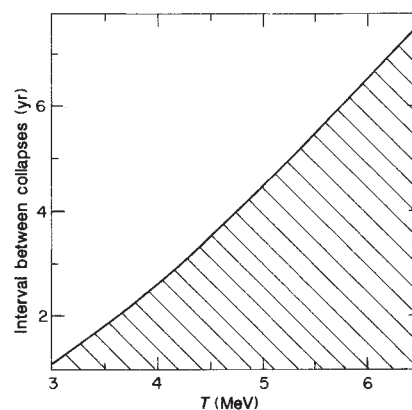


Fig. 3 Limits that might be imposed on the mean interval between galactic stellar collapses by the ^{97}Mo experiment as a function of the mean ν_e temperature. The shaded area would be ruled out. A fixed average energy in ν_e s of 0.8×10^{53} erg per collapse and a solar neutrino ^{97}Tc production rate of 3.9×10^{-36} per s per target atom are assumed. (The latter would be given by a low-metallicity 'consistent' solar model.)

$^{97}\text{Mo}(\nu_e, e^-)^{97}\text{Tc}$ are 1.4 SNU and 0.17 SNU, respectively, for a total of 40% of the solar signal.

It follows that there could be a sizeable background in the ^{97}Tc experiment because of galactic neutrinos. Conversely, the experiment may yield an interesting limit on the frequency and energy release of galactic stellar collapses. The most conservative strategy would be to derive such a limit by attributing the entire ^{97}Tc production to galactic neutrinos. Collapse frequencies $\geq 0.25 \text{ yr}^{-1}$ could be ruled out, provided uranium-associated backgrounds in the ^{97}Tc experiment prove tolerable. Alternatively, to the extent that the rate of solar neutrino production of ^{97}Tc can be determined, this limit could be tightened. The solar neutrino production rate of ^{98}Tc might be useful in such a subtraction.

The results that follow from the first strategy are given by the solid line in Fig. 3. The calculation assumes that a ^{97}Tc concentration is found consistent with the low-metallicity model. Limits on the frequency of stellar collapses are derived by adopting an effective average energy in ν_e s of 0.8×10^{53} erg per collapse, but allowing the average neutrino temperature to vary. The sharp growth in the cross-section with increasing mean E_ν leads to more stringent limits on the stellar collapse frequency for large T . The relation is approximately linear (frequency $\times T \approx \text{constant}$).

In conclusion, we emphasize that high-threshold continuum nuclear reactions can be exploited both to maximize the nuclear response to galactic neutrinos and to minimize backgrounds from solar neutrinos. As capture rates on the order of one SNU could be achieved for some (ν_e, e^-) and neutral current reactions, geochemical measurements of the long-term galactic neutrino flux would seem to be possible, in principle. The number of potentially practical experiments is limited. However, as such measurements could provide a unique perspective on the recent evolution of our galaxy, potential experiments deserve serious consideration.

Received 12 November 1987; accepted 4 March 1988.

1. Scott, R. D. *Nature* **264**, 729-730 (1976).
2. Freedman, M. S. *et al. Science* **193**, 1117-1119 (1976).
3. Rowley, J. K., Cleveland, B. T., Davis, R., Hampel, W. & Kirsten, T. in *The Ancient Sun* (eds Pepin, R. O., Eddy, J. A. & Merrill, R. B.) 45-62 (Pergamon, New York, 1980).
4. Cowan, G. A. & Haxton, W. C. *Science* **216**, 51-54 (1982).
5. Hirata, K. *et al. Phys. Rev. Lett.* **58**, 1490-1493 (1987).
6. Bionta, R. *et al. Phys. Rev. Lett.* **58**, 1494-1496 (1987).
7. Woosley, S. E. & Weaver, T. A. *Ann. Rev. Astr. Astrophys.* **24**, 205-253 (1986).
8. Wolfsberg, K. *et al. in Solar Neutrinos and Neutrino Astronomy* (eds Cherry, M. L., Fowler, W. A. & Lande, K.) 196-202 (American Institute of Physics, New York, 1985).
9. Wilson, J. R. in *Numerical Astrophysics* (eds Centrella, J., LeBlanc, J. & Bowers, R.) 422-434 (Jones & Bartlett, Boston, 1986).
10. Bahcall, J. N., Piran, T., Press, W. H. & Spergel, D. *Nature* **327**, 682-685 (1987).
11. Krauss, L. *Nature* **329**, 689-694 (1987).

12. Schramm, D. N. *Comments Nucl. Part. Phys.* **17**, 239-278 (1987).
13. Kahana, S. H., Cooperstein, J. & Baron, E. *Phys. Lett.* **B196**, 259-266 (1987).
14. Wheeler, J. C. & Wheeler, J. A. in *Science Underground* (eds Nieto, M. M. *et al.*) 214-224 (American Institute of Physics, New York, 1983).
15. Bahcall, J. N. & Piran, T. *Astrophys. J.* **267**, L77-L81 (1983).
16. Tkaczyk, W., Karakula, S. & Giovanelli, F. *Nuovo Cim.* **9C**, 435-442 (1986).
17. Bahcall, J. N. & Soneira, R. M. *Astrophys. J. Suppl.* **44**, 73-110 (1980).
18. Terry, K. D. & Tucker, W. H. *Science* **159**, 421-423 (1968).
19. Rapaport, J. *et al. Phys. Rev. Lett.* **54**, 2325-2328 (1985).
20. Rowley, J. K., Cleveland, B. T. & Davis, R., Jr. in *Solar Neutrinos and Neutrino Astronomy* (eds Cherry, M. L., Fowler, W. A. & Lande, K.) 1-21 (American Institute of Physics, New York, 1985).
21. Bahcall, J. N. & Ulrich, R. K. *Rev. Mod. Phys.* (in the press).
22. Mikheyev, S. P. & Smirnov, A. Yu. *Sov. J. Nucl. Phys.* **42**, 913-917 (1985).
23. Mikheyev, S. P. & Smirnov, A. Yu. *Nuovo Cim.* **9C**, 17-26 (1986).
24. Krauss, L. M., Glashow, S. L. & Schramm, D. N. *Nature* **310**, 191-198 (1984).
25. Kuo, T. T. S. & Brown, G. E. *Nucl. Phys.* **A114**, 241-279 (1968).

Homologous plant and bacterial proteins chaperone oligomeric protein assembly

Sean M. Hemmingsen*, Carol Woolford†, Saskia M. van der Vies‡, Kit Tilly§, David T. Dennis||, Costa P. Georgopoulos¶, Roger W. Hendrix# & R. John Ellis‡

* Plant Biotechnology Institute, National Research Council, Saskatoon, Saskatchewan, S7N 0W9, Canada

† Department of Biological Sciences, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA

‡ Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK

§ Laboratory of Biochemistry, NCI-NIH, Bethesda, Maryland 20892, USA

|| Department of Biology, Queen's University, Kingston, Ontario, K7L3N6, Canada

¶ Department of Cellular, Viral, and Molecular Biology, University of Utah Medical Center, Salt Lake City, Utah 84132, USA

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

An abundant chloroplast protein is implicated in the assembly of the oligomeric enzyme ribulose biphosphate carboxylase-oxygenase, which catalyses photosynthetic CO₂-fixation in higher plants. The product of the Escherichia coli groEL gene is essential for cell viability and is required for the assembly of bacteriophage capsids. Sequencing of the groEL gene and the complementary cDNA encoding the chloroplast protein has revealed that these proteins are evolutionary homologues which we term 'chaperonins'. Chaperonins comprise a class of molecular chaperones that are found in chloroplasts, mitochondria and prokaryotes. Assisted post-translational assembly of oligomeric protein structures is emerging as a general cellular phenomenon.

THE process of post-translational assembly of polypeptides into oligomeric structures is an often overlooked aspect of the molecular biology of gene expression. The assembly process is generally assumed to result from properties inherent in the structures of the polypeptide subunits themselves, because several striking examples of this principle of self-assembly are known¹. In at least some cases, however, structural information resident in other protein molecules is required for the correct assembly of oligomeric proteins. These assembly proteins have been termed 'molecular chaperones' because they prevent the formation of incorrect structures but are not part of the final oligomer².

Correct post-translational assembly of the head proteins of bacteriophages lambda and T4, and of the tail proteins of bacteriophage T5, requires the function of the *Escherichia coli* *groEL* gene product³⁻⁶. A protein in the stroma of higher plant chloroplasts, termed the Rubisco subunit binding protein (referred to later simply as 'binding protein'), is implicated in the post-translational assembly of the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco)^{7,8}. We report here that, based on the similarity of their predicted amino-acid sequences, the *groEL* gene product and the Rubisco subunit binding protein are evolutionary homologues. Data presented here, and additional data in press, indicate that the proteins of this highly conserved class are ubiquitous, or nearly so. The conservation and ubiquity of this class of proteins indicate that its specific roles in the assembly of Rubisco and of bacteriophage structural proteins reflect a more general and essential cellular function. We propose that this class of proteins be termed 'chaperonins'.

The *groE* gene products

The *E. coli* *groE* locus was first identified in studies of mutations that interfere with the assembly of the heads of bacteriophages lambda and T4 (refs 3, 4, 9, 10). It was subsequently shown that the *groE* products are also essential for bacteriophage T5 tail assembly⁶. The *groE* locus maps to 94 min on the *E. coli* genetic map¹¹ and is composed of two genes, *groEL* and *groES*¹², both of which are essential for cell viability (Fayet, O, Ziegelhoffer, T., C.P.G., personal communication). The *groEL* and *groES* genes encode polypeptides with apparent relative molecular masses (*M_r*) of 65,000 and 15,000 respectively^{12,13}, which are among the most abundant proteins in the cell. Both

the *groEL* and *groES* polypeptides are found in the cell as oligomeric structures. The purified *groEL* protein is a 14 subunit homo-oligomer composed of two stacked rings of 7 subunits each^{14,15}. The *groES* protein consists of 6-8 identical subunits arranged in a single ring¹⁶. The *groEL* protein is an ATPase¹⁴, and the *groEL* and *groES* proteins form a complex with each other in the presence of Mg-ATP but not in its absence¹⁶. The *groE* proteins act at an early stage in the head assembly pathways of bacteriophages lambda, T4, and others¹⁷. In the lambda head assembly pathway, both *groE* proteins are required for the earliest assembly step that has been detected, which is the formation of an oligomer of 12 subunits of phage protein gpB⁵; this oligomer subsequently forms the 'connector' or 'portal' structure on which head-shell assembly occurs. In the absence of *groEL* function, the T4 major head protein is found associated with the cell membrane in amorphous 'lumps'⁹. Although the *groE* proteins participate in normal phage head assembly, they are not found in the assembled head structures¹⁸.

The *groE* genes are members of the *E. coli* heat shock regulon. Synthesis of *groEL* protein, which accounts for 1% of cellular protein synthesis in cells in steady state growth at 37 °C, increases to 10% of total synthesis soon after the cells are shifted to a growth temperature of 46 °C. The promoter controlling transcription of the *groE* genes is recognized by RNA polymerase carrying the heat shock sigma factor, δ 32 (ref. 20).

The DNA sequence of the *groE* operon is presented in Fig. 1. Examination of this sequence shows two open reading frames encoding polypeptides of *M_r* 10,368 and *M_r* 57,259 and predicted *pI* values of 5.92 and 5.63 respectively. These estimates correspond with the respective measured values for the *groES* and *groEL* polypeptides. Comparison of amino-acid compositions and amino-terminal sequences determined from the purified *groES* and *groEL* proteins with the compositions and sequences deduced from the DNA sequence show excellent agreement. Northern blot experiments (not shown) reveal a single transcript of ~2,100 nucleotides containing both *groES* and *groEL* sequences. This transcript, which is four to five times more abundant in extracts of heat-shocked cells compared to non-heat-shocked cells, extends from the previously identified promoter upstream of *groES* to a predicted terminator stem-loop structure ~60 nucleotides 3' from the *groEL* coding region. The *groEL* and *groES* genes are therefore co-transcribed as an operon.



Fig. 1 Nucleotide sequence of the *E. coli* *groE* operon. The predicted amino-acid sequences of the *groES* and *groEL* open reading frames are indicated, as well as the promoter and transcription start site²⁰, and the dyad symmetry of the proposed transcription terminator. The predicted amino-terminal sequences match the chemically determined sequences of the *groES* and *groEL* proteins. (For *groES*, 8 of 8 residues match, including the amino-terminal methionine (data not shown); 21 of 23 residues match for *groEL*⁴⁷. The amino-terminal methionine is absent in mature *groEL* protein.) The source of DNA for sequencing was the 7.7 kilobase (kb) insert in the *groE* transducing phage H18 (ref. 13). Parts of this DNA were subcloned into plasmid pBR322 and the sequence was determined either directly by the method of Maxam and Gilbert⁴⁸ (residues 1-956 and 1953-2198) or by the method of Sanger *et al.*⁴⁹ following further subcloning into phage M13 vectors (residues 644-2267). The sequences of both strands of the DNA were determined, and all restriction sites were crossed.

The Rubisco subunit binding protein

The enzyme Rubisco occurs in higher plant chloroplasts as an oligomer of eight large subunits, which are synthesized within the chloroplast, and eight small subunits, which are synthesized by cytosolic ribosomes². The Rubisco subunit binding protein was first identified in studies with isolated intact chloroplasts of *Pisum sativum* (pea) as a protein that binds newly synthesized Rubisco large subunits before their appearance in the holoenzyme⁷. The binding protein also non-covalently binds newly synthesized Rubisco small subunits after their import from the cytosol⁸. The binding protein is not itself a component of the assembled Rubisco enzyme.

The binding protein consists of two types of subunit of apparent M_r 61,000 (alpha) and M_r 60,000 (beta)²¹. The subunits are distinct by a number of criteria²¹, but amino-terminal amino-acid sequence analysis shows some sequence similarity²². The subunits are encoded by nuclear genes and are imported into the chloroplast after synthesis in the precursor form by cytosolic ribosomes²³. The purified oligomeric binding protein has an apparent M_r estimated by gel-sieving to be greater than 700,000 (ref. 23). The binding protein oligomer dissociates reversibly *in vitro* in the presence of Mg-ATP and can be recovered in forms ranging from the high M_r forms to monomers^{21,23-25}. The purified protein has been reported to show ATPase activity²⁶.

There is evidence that the binding protein plays an essential role in the assembly of Rubisco in higher plants⁸. The most

convincing evidence is the inhibition of the transfer of newly synthesized Rubisco large subunits to the holoenzyme in stromal extracts by the addition of antiserum to the binding protein^{27,28}. In contrast to the higher plant Rubisco, purified Rubisco from cyanobacteria can be dissociated into its subunits and subsequently reassembled into an active enzyme *in vitro*²⁹. An active enzyme is also produced when the cyanobacterial Rubisco genes are expressed in *E. coli*³⁰. Efforts in several laboratories to extend these results to higher plant Rubisco have met with failure however³⁰. The large subunit of higher plant Rubisco is apparently incapable of participating in assembly either after dissociation from the enzyme or after heterologous expression in the absence of the binding protein^{29,30}. The failure of assembly in *E. coli* is blocking attempts to produce improved forms of this agriculturally important enzyme by *in vitro* mutagenesis⁸.

Clones encoding castor-bean (*Ricinus communis*) and wheat (*Triticum aestivum*) binding proteins were isolated from cDNA libraries constructed in the expression vector λ gt11 and screened with a specific polyclonal antibody raised against the binding protein from *P. sativum* leaf chloroplasts²³. The sequences of the plant cDNAs and the predicted amino-acid sequences are presented in Fig. 2.

The wheat cDNA sequence contains an open reading frame encoding a polypeptide with a predicted pI of 5.63. The predicted amino-terminal sequence corresponds well with that determined for the amino-terminus of the purified mature alpha

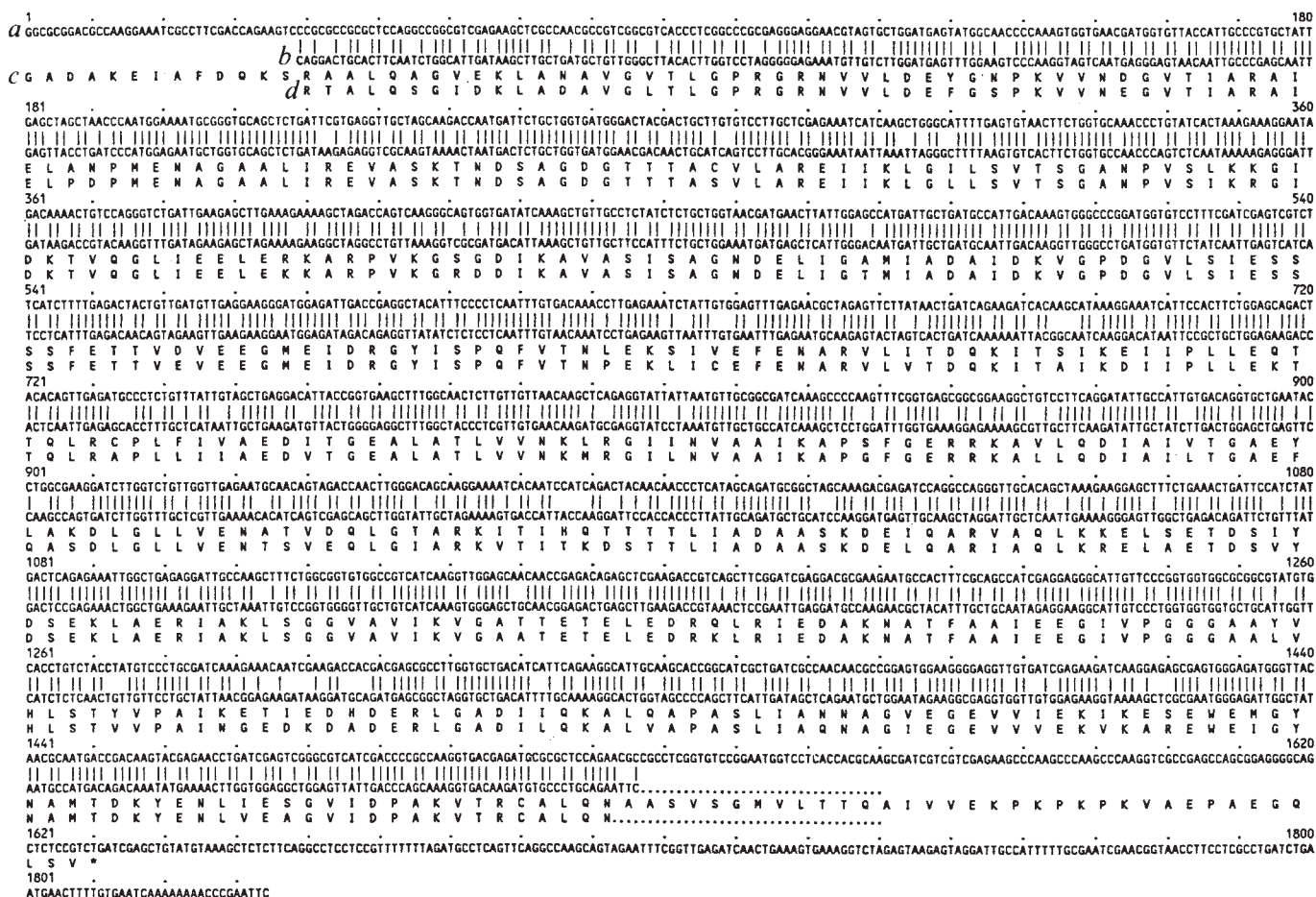


Fig. 2 Nucleotide sequences of the *R. communis* and *T. aestivum* Rubisco subunit binding-protein alpha subunits. The *T. aestivum* (a) and *R. communis* (b) nucleotide sequences and the predicted amino-acid sequences of the open reading frames are indicated (*T. aestivum*, (c) and *R. communis*, (d)). Vertical lines indicate nucleotide identity.

Methods. Libraries of cDNA derived from poly(A⁺) RNA extracted from the endosperm of developing *R. communis* seeds and from young wheat leaves, were used to construct λ gt11 expression libraries⁵⁰. The *T. aestivum* library was kindly provided by C. Raines and T. A. Dyer. The *R. communis* library was made without methylation of the cDNA before *EcoRI* digestion and insertion into the vector. Both libraries were screened with a specific polyclonal antibody raised against purified binding protein from *P. sativum* leaf chloroplasts²³. From 150,000 *R. communis* library recombinants screened, 26 positives were recovered. The inserts from four of these were subcloned into a plasmid vector for further analysis. Each of the subclones hybridized specifically to a transcript of 2,300 bp in total *R. communis* RNA. Sequence determination of the *R. communis* clones was by primer extension using the chain termination method of Sanger *et al.*⁴⁹. A recombinant isolated from the *T. aestivum* expression library contained a cDNA insert of ~2 kb, which hybridizes very strongly to the isolated *R. communis* cDNAs. This fragment was subcloned in the plasmid vector PUBS (a derivative of Bluescript kindly provided to us by G. Murphy) giving rise to pSV10 and pSV11 (opposite orientation); pSV10 and pSV11 were used to determine the *T. aestivum* nucleotide sequence⁵¹.

polypeptide of the binding protein from both *T. aestivum* (17 matches out of 20) and *P. sativum* (13 matches out of 20). There is less correspondence between the predicted wheat sequence and the amino-terminal sequence of the purified mature beta subunit of *P. sativum* (9 matches out of 20) and four of these matches are seen only if account is taken of two extra residues in the pea beta subunit (Fig. 3). We conclude that the wheat cDNA clone encodes the entire mature binding protein alpha subunit plus two amino acid residues of the presequence. The calculated M_r of the mature alpha subunit is 57,393. The castor-bean cDNA sequence contains an internal 3' *EcoRI* site and it does not extend to the mature amino-terminus corresponding to that determined for the *P. sativum* alpha subunit; it does, however, include 91% of the coding sequence for the mature polypeptide based on the wheat sequence. The wheat and castor-bean sequences are 80% identical at the amino-acid level.

Similarity of plant and bacterial sequences

The predicted amino-acid sequences of the wheat binding-protein alpha subunit and the *groEL* protein are compared in

Fig. 4. The overall amino-acid compositions are very similar and the calculated M_r and pI values are almost identical. With the introduction of only a few gaps, 46% of the residues are identical and many of the differences represent conservative substitutions. On the basis of the high sequence similarity, we conclude that the binding and *groEL* proteins are homologues.

An earlier report showed that total leaf extracts of *P. sativum* contain a protein similar to the *groEL* protein in that it consists of 14 subunits, arranged in two layers of 7 monomers each, and shows weak ATPase activity³¹. It seems likely that the protein these authors observed is identical with the Rubisco subunit binding protein. We suggest that the binding protein consists of seven alpha and seven beta subunits, and this extends the similarity between the *groEL* protein and the binding protein to the level of ultrastructure.

The plant and bacterial proteins both show weak ATPase activities and the oligomeric structures of both are affected by the presence of Mg-ATP. The oligomeric structures may be stabilized by hydrophobic interactions upon which the Mg-ATP acts. The dissociation of the plant protein is enhanced at lower

a DAKEIAFDQK SRAALQAGVE
 * * * * *
 b XAPEIAFDQG SRAALQAGVE
 a DAKEIAFDQK SRAALQAGVE
 ** * * * *
 c AAKDIAFDQH SRSAMQAGID
 a AKEIAFDQKS RAA..LQAGV
 *** * * * *
 d AKELHFNKDG SAIRKLQNGV

Fig. 3 Comparison of the amino-terminal sequences of the binding-protein subunits from *T. aestivum* and *P. sativum*. A lyophilized sample of the *T. aestivum* binding-protein alpha subunit (kindly provided by A. J. Keys) was subjected to automated solid-phase Edman degradation by the SERC protein sequencing unit at the Department of Biochemistry, University of Leeds. Symbols: a, predicted sequence of cDNA from *T. aestivum*; b, determined sequence of binding-protein alpha subunit from *T. aestivum*; c, determined sequence of binding-protein alpha subunit from *P. sativum*²¹; d, determined sequence of binding-protein beta subunit from *P. sativum*²¹. Asterisks indicate identical residues, dots indicate gaps inserted to maximize the identity. The three discrepancies between the determined sequence and the predicted sequence of the *T. aestivum* subunit (lines a and b) may result from incomplete recovery of aminoterminal and lysine residues associated with the sequencing method; alternatively, they may reflect varietal differences, because the alpha subunit was obtained from variety 'Ficco' while the cDNA was prepared from variety 'Chinese Spring'.

temperature²³ and the bacterial protein will dissociate in metrizamide, which is less polar than water, in the presence of Mg-ATP (Chandrasekhar, G. N. & C.P.G., personal communication). The requirement for Mg-ATP for the association of the *groEL* and *groES* proteins may in fact be a requirement for the destabilisation of the interactions of *groEL* monomers amongst themselves. These similarities suggest that the proteins retain similar physical and functional properties throughout evolution.

Ubiquitous occurrence

Bacterial DNA sequences with a high degree of similarity to the *groEL* coding sequence have been identified in a number of other bacterial species³². The *groEL*-like protein is a major immunogen in a number of infections including those caused by mycobacteria. Studies using antibodies to the antigen from *Mycobacterium leprae* and *M. tuberculosis* have shown that this antigen is extremely widespread in prokaryotes, including archaeobacteria^{33,34}. Antibodies to the *P. sativum* binding protein have been used independently and recognize a polypeptide of apparent M_r 60,000 in similarly diverse groups of bacteria (S.M.H., S.M.V. & R.J.E., unpublished observations).

An abundant protein with an apparent subunit M_r of 58,000 has been purified recently from *Tetrahymena thermophila* and shown to be related to the *groEL* protein both immunologically and structurally^{35,36}. This *groEL*-like protein is localized in the mitochondria, where its suggested function is to assist in the refolding of newly imported proteins. Antibodies to the *Tetrahymena groEL*-like protein cross-react with a protein of apparent subunit M_r of 58–64,000 in extracts of mitochondria from yeast, *Xenopus*, maize and human, suggesting that it occurs in all mitochondria³⁶.

Of the three bacterial strains with sequenced *groEL* homologues, one, *Coxiella burnetii*, has a homologue of the *groES* gene, which, as in the *E. coli* case, lies a short distance upstream from the *groEL* gene (Vodkin, M. & Williams, J., personal communication). The *C. burnetii groES* protein shows 53% amino-acid sequence identity with the *E. coli groES* protein. For the two mycobacterial species, there is no evidence for sequence matches to the *groES* gene, despite the availability of several hundred base pairs (bp) of sequence on both sides of the *groEL* coding regions. There is no information on the possible occurrence of the *groES* sequence in eukaryotes.

a 1 GADAKEIAFDQKSRAALQAGVEKLANAVGVTGPRGRNVVLDE.YGNPKV 49
 b 1 .MAAKDVKFGNDARVKMLRGVNVVLADAVKVTGPKGRNVVLDSKFGAPT 49
 50 VNDGVTIARAIELANPMENAGAALIREVASKTNDSDAGDTTACVLAREI 99
 50 TKDGVSVAREIELEDKFENMGAQMVKVEASKANDAAGDGNNTATVLAQAI 99
 100 IKLGILSVTSGANPVSLKKGIDKTQGLIEELERKARPVKSGSDIKAVAS 149
 100 ITEGLKAVAAGMNPMDLKRIGDKAVTAAVEELKALSVPCSDSKAIAQVGT 149
 150 ISAGNDELIGAMIADAIDKVGPDGVLSIESSSSFETTVDDVEEGMEIDRGY 199
 150 ISATSDETVGKLIAEAMDVKVKEGVITVEDGTGLQDELVDVVEGMQFDRGY 199
 200 ISPQFVTNLEKSIVEFENARVLITDQKITSIKEIIPLEQTTLQRCPLFI 249
 200 LSPYFINKPETGAVELESFILLADKKISNIREMLPVEAVAKAGKPLLI 249
 250 VAEDITGEALATLVVNKLRLGIINVAIAKPSFGERRKAVLQDIAIVTGA 299
 250 IATEDVEGEALATAVNTIRGIVKVAVKAPGFGDRRKAMLDIATLTGGT 299
 300 YLAKDLGLLVENATVDQLGTARKITIHQTTLTIADAASKDEIQARVAQL 349
 300 VISEEIGMELEKATLEDLGQAKRVVINKDITTIIDGVGEAAIQGRVAQI 349
 350 KKELETSDSIYDSEKLAERIAKLSGGVAVIKVGATTETEDRLRLIEDA 399
 350 RQIEEATSDYDREKLQERVAKLGGVAVIKVGAATEVEMKEKKARVEDA 399
 400 KNATFAAIEEGIVPGGGAAYVHLSTYVPAIKETIEDHDERLGDIIQKAL 449
 400 LHATRAAVEEGVAGGVALIRVASKLADLRG..QNEDQNVGIKVALRAM 447
 450 QAPASLIANNAGVEGEVVIKIKESWEMGYNAMTDKYENLIESGVIDPA 499
 448 EAPLRQIVLNCGEPSVAVANTVKGGDGNYGNAATEEYGNMIDMGILPT 497
 500 KVTICALQNAASVSGMVLTTQAIIVVEKPKPKPKVAEPAGQLSV* 544
 498 KVTISALQYASVAGLMITTECMVTDLPKNDADLGAAGMGGMGGMM* 549

Fig. 4 Predicted amino-acid sequences of the wheat binding protein alpha subunit (a) and the *E. coli groEL* protein (b). Vertical lines indicate amino-acid identity, and dots indicate gaps inserted to maximize identity. The first two residues of the wheat sequence are the last two residues of the presequence.

Discussion

We have described a ubiquitous, conserved, abundant protein that is associated with the post-translational assembly of at least two structurally distinct oligomeric protein complexes. We conclude that the role of this protein is to assist other polypeptides to maintain or to assume conformations which permit their correct assembly into oligomeric structures. We propose that this protein be called a 'chaperonin', since its properties meet the criteria suggested for molecular chaperones². The term 'molecular chaperone' was first used to describe nucleoplasmin, an acidic nuclear protein required for the assembly of nucleosome cores from DNA and histones in extracts of *Xenopus* eggs³⁷. The term has now been extended to describe a family of proteins whose proposed role is to ensure that the folding of certain polypeptides and their assembly into oligomeric structures occur correctly². It has been suggested that the essential function of molecular chaperones is to prevent the formation of 'improper' structures which might result from the transient exposure of hydrophobic or charged surfaces normally involved in domain interactions, either within or between polypeptide chains. Such exposure may occur during the operation of several cellular processes. These include the synthesis of polypeptides, the refolding that occurs after their transport across membranes, the association of polypeptides made in one subcellular compartment with those made in another, changes in protein-protein interactions during the normal functioning of a complex and recovery from stress such as heat shock. The other criterion that distinguishes molecular chaperones is that they do not form part of the final structure whose assembly they have mediated. We recognize currently three classes within the molecular chaperone family, namely, (1) nucleoplasmin, (2) *hsp70*-immunoglobulin heavy chain binding protein class, and (3) the bacterial-mitochondrial-chloroplast class. We propose to limit the term 'chaperonin' to the last class.

Some of the chaperonins in both prokaryotes and eukaryotes are known to be heat-shock proteins^{19,35,36}. It has been proposed recently that heat-shock proteins may function both to refold heat-denatured proteins and to influence the folding or aggregation state of other proteins under non-stress conditions³⁸. This proposal supports the functions we have suggested for the molecular chaperones.

Some of the proteins most highly conserved between prokaryotes and eukaryotes are found associated with mitochondria and chloroplasts, for example the beta subunit of ATP synthetase³⁹ and elongation factor EFTu⁴⁰. As with these proteins, the occurrence of the highly conserved chaperonins in both mitochondria and chloroplasts, as well as in prokaryotes, is most easily explained in terms of the endosymbiotic hypothesis for the origin of these organelles. We suggest that the original role of prokaryotic chaperonin has expanded during evolution to include the mediation of processes peculiar to the biogenesis of these organelles, especially the refolding of polypeptides imported from the cytosol and their combination with polypeptides synthesized within the organelle. We note that the high level of sequence identity that we report here between the prokaryotic and eukaryotic chaperonins is very similar to the levels of identity reported elsewhere between the prokaryotic and eukaryotic members of the *hsp70* and *hsp83* heat-shock protein classes^{41,42}. It thus seems possible that the eukaryotic heat-shock proteins as a class were derived from prokaryotes endosymbiotically, even though the *hsp70* and *hsp83* proteins are not now located in organelles.

It remains to consider what the normal roles of chaperonin might be in prokaryotes in the absence of phage infection or environmental stresses. Deletion mutations of the *groE* operon cannot be isolated at temperatures ranging from 20°C to 43°C, demonstrating that the *groE* proteins are essential for *E. coli* growth at all temperatures (Fayet, O., Ziegelhoffer, T., C.P.G., personal communication). Mutations in the *groEL* and *groES* genes interfere with DNA and RNA synthesis at high temperatures⁴³, while overproduction of both the *groEL* and *groES* proteins from a plasmid suppresses certain mutations in the *dnaA* gene, which is required for the initiation of DNA replication^{44,45}. Similarly, the presence of a plasmid carrying what now appears to be the *groE* operon (or even one carrying an internal fragment of the *groEL* gene) suppresses a temperature-sensitive mutation in *ams*, an *E. coli* gene that influences messenger RNA half-life⁴⁶. One interpretation of these observations is that

the *groEL* and *groES* chaperonins function normally to mediate protein-protein interactions involved in nucleic acid metabolism in *E. coli*; a similar suggestion has been made to explain the sequence similarity between the major heat shock proteins of animal cells (*hsp70*) and the product of the *dnaK* gene in *E. coli*³⁸. It is also conceivable that the chloroplast and mitochondrial chaperonins function in DNA replication within these organelles and that the chaperonins mediate the assembly of oligomeric complexes other than those involved in nucleic-acid metabolism. Future work will aim to define the scope of chaperonin action in the formation of protein complexes in both prokaryotes and eukaryotes. The idea that molecular chaperones are involved transiently in the folding or assembly of certain oligomeric proteins also has implications for those biotechnologists who are encountering problems in producing foreign proteins in bacterial cells in the required active form. It may be necessary to re-examine the natural synthesis of these proteins by techniques which permit the detection of noncovalently bound complexes to determine whether molecular chaperones are involved.

The involvement of the chloroplast chaperonin in the assembly of the higher plant Rubisco could account for the failure to obtain the synthesis of enzymatically active Rubisco oligomers in *E. coli*³⁰. It would follow that the bacterial chaperonin must be sufficiently different from the chloroplast chaperonin to be unable to mediate the assembly of the Rubisco from higher plants. The observation that Rubisco from cyanobacteria can be synthesized in active oligomeric form by *E. coli*³⁰ suggests either that the *E. coli* chaperonin is able to mediate the assembly of the Rubisco from prokaryotes or that the assembly of Rubisco in cyanobacteria does not involve chaperonin. Our current aim is to express the cDNAs for the chloroplast chaperonin in the same *E. coli* cells that are expressing the genes for Rubisco from higher plants, in the hope of rescuing Rubisco assembly and so permitting attempts to improve the properties of this agriculturally important enzyme.

We thank D. Bradley, V. Knowles, M. A. McCrae and G. Murphy for assistance with DNA sequencing, J. Keen for determination of the amino-terminal protein sequence, C. Raines and T. A. Dyer for access to a wheat cDNA library, H. F. Kung, D. B. Young and R. A. Young for helpful discussion and SERC and NIH for financial support. The sequence data in this publication are being submitted to the EMBL/GenBank Libraries.

Received 14 March; accepted 20 April 1988.

- Alberts, B. et al. *Molecular Biology of the Cell* (Garland, New York, 1983).
- Ellis, R. J. *Nature* **328**, 378-379 (1987).
- Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R. & Kaiser, A. D. *J. molec. Biol.* **76**, 45-60 (1973).
- Sternberg, N. *J. molec. Biol.* **76**, 25-44 (1973).
- Kochan, J. & Murialdo, H. *Virology* **131**, 100-115 (1983).
- Zweig, M. & Cummings, D. J. *J. molec. Biol.* **80**, 505-518 (1973).
- Barracough, R. & Ellis, R. J. *Biochim. Biophys. Acta* **608**, 19-31 (1980).
- Ellis, R. J. & van der Vies, S. M. *Photosynthesis Res.* (in the press).
- Takano, T. & Kakefunda, T. *Nature new Biol.* **239**, 34-37 (1972).
- Coppo, A., Manzi, A., Pulitzer, J. F. & Takahashi, H. *J. molec. Biol.* **76**, 61-87 (1973).
- Georgopoulos, C. P. & Eisen, H. J. *supramolec. Struct.* **2**, 349-359 (1974).
- Tilly, K., Murialdo, H. & Georgopoulos, C. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1629-1633 (1981).
- Georgopoulos, C. P. & Hohn, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 131-135 (1978).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375-392 (1979).
- Hohn, T., Hohn, B., Engel, A., Wurtz, M. & Smith, P. R. *J. molec. Biol.* **129**, 359-373 (1979).
- Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. J. *biochem. Chem.* **261**, 12414-12419 (1986).
- Georgopoulos, C. P., Tilly, K. & Casjens, S. R. in *Lambda II* (eds Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 279-304 (Cold Spring Harbor Monograph 13, Cold Spring Harbor, 1983).
- Friedman, D. I. et al. *Microbiol. Rev.* **48**, 299-325 (1984).
- Neidhardt, F. C. & VanBogelen, R. A. in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 1334-1345 (American Society for Microbiology, 1987).
- Cowing, D. W. et al. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2679-2683 (1986).
- Musgrove, J. E., Johnson, R. A. & Ellis, R. J. *Eur. J. Biochem.* **163**, 529-534 (1987).
- Ellis, R. J., van der Vies, S. M. & Hemmingsen, S. M. in *Plant Molecular Biology 1987* (eds von Wettstein, D. & Chua, N.-H.) 33-40 (NATO-ASI Series A, Life Sciences Vol. 140, Plenum, New York and London, 1987).
- Hemmingsen, S. M. & Ellis, R. J. *Pl. Physiol.* **80**, 269-276 (1986).
- Musgrove, J. E. & Ellis, R. J. *Phil. Trans. R. Soc. B313*, 419-428 (1986).
- Bloom, M. V., Milos, P. & Roy, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1013-1017 (1983).
- Chaudhari, P., Cannon, S., Hubbs, A. & Roy, H. Abstracts 16th Annual UCLA Symposium, *J. cell Biochem. suppl.* **11B**, 50 (1987).
- Milos, P. & Roy, H. *J. cell Biochem.* **24**, 153-162 (1984).
- Cannon, S., Wang, P. & Roy, H. *J. Cell Biol.* **103**, 1327-1335 (1986).
- Andrews, T. J. & Lorimer, G. H. *J. biochem. Chem.* **260**, 4632-4636 (1985).
- Bradley, D., van der Vies, S. M. & Gatenby, A. A. *Phil. Trans. R. Soc. B313*, 447-458 (1986).
- Pushkin, A. V. et al. *Biochim. biophys. Acta* **704**, 379-384 (1982).
- Young, D., Lathigra, R., Hendrix, R., Sweetser, D. & Young, R. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Thole, J. E. R. *Microbiol. Pathogenesis* (in the press).
- Young, D. B., Ivanyi, J., Cox, J. H. & Lamb, J. R. *Immunology Today* **8**, 215-219 (1987).
- McMullin, T. W. & Hallberg, R. L. *Molec. cell Biol.* **7**, 4414-4423 (1987).
- McMullin, T. W. & Hallberg, R. L. *Molec. cell Biol.* **8**, 371-380 (1988).
- Laskey, R. A., Honda, B. M., Mills, A. P. & Finch, J. T. *Nature* **275**, 416-420 (1978).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Zurawski, G. & Clegg, M. T. *Nucleic Acids Res.* **12**, 2549-2559 (1984).
- Nagata, S., Tsunetsugu-Yokota, Y., Naito, A. & Kaziro, Y. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6192-6196 (1983).
- Bardwell, J. & Craig, E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 848-852 (1984).
- Bardwell, J. & Craig, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5177-5181 (1987).
- Wada, M. & Itikawa, H. *J. Bact.* **157**, 694-696 (1984).
- Fayet, O., Louarn, J.-M. & Georgopoulos, C. *Molec. gen. Genet.* **202**, 435-445 (1986).
- Jenkins, A. J., March, J. B., Oliver, I. R. & Masters, M. *Molec. gen. Genet.* **202**, 446-454 (1986).
- Chanda, P. K., Ono, M., Kuwano, M. & Kung, H.-F. *J. Bact.* **161**, 446-449 (1985).
- Subramanian, A. R., Wittman-Liebold, B., Geissler, A. W., Stoffler, G. & Giesen, M. *FEBS Lett.* **99**, 357-360 (1979).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 449-560 (1980).
- Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194-1198 (1983).
- Murphy, G. & Kavanagh, T. *Nucleic Acids Res.* (in the press).

The fate of cluster gas in neutrino-dominated cosmologies

August E. Evrard* & Marc Davis†

* Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

† Departments of Astronomy and Physics, University of California, Berkeley, California 94720, USA

Once popular several years ago, cosmologies with massive neutrinos as the dominant mass component¹⁻⁴ were shown to suffer from several serious deficiencies. Foremost among the problems was the large clustering length predicted for matter and for galaxies, if they formed in regions of caustic surfaces⁵⁻⁸. It was argued that this problem could be circumvented if galaxies formed preferentially in low density regions of space, that is, if the galaxy distribution were anti-biased relative to the mass⁹⁻¹⁰. But further study¹¹ showed that even if a working form of anti-biasing were found to alleviate the 'length scale problem' of galaxy formation, hot gas confined within the rich neutrino clusters would produce copious X-ray emission. The expected sizes, number density, temperatures and luminosities of these clusters would be in disagreement with observational data, unless only a very small fraction of the universe was in baryons, $\Omega_b \lesssim 0.025$, and the universe was quite young, $h \approx 1$ where $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$.

A possible solution to this second, 'over-rich cluster' problem would be to require that baryons somehow be kept out of cluster cores. Prior heating by shocks or feedback from an early generation of radiating sources have been suggested as possible mechanisms^{9,17}. The hope is to increase substantially the entropy in the gas before cluster formation, preventing it from being compressed to high densities during cluster formation. Because the core X-ray luminosity, L_X scales with pressure P and baryon density ρ_b as

$$L_X \propto \rho_b^{3/2} P^{1/2} \quad (1)$$

maintaining constant pressure at a lower density (increasing the entropy) can substantially decrease the X-ray emission from a cluster.

Here we show, by simulation methods, that shock heating before cluster formation cannot do this. The gas traces the underlying neutrino distribution in both density and temperature. A strong positive entropy gradient is found for gas in clusters; central gas is shocked weakly compared to material falling later into the cluster. We argue that conventional sources of feedback operating after the core collapse of a cluster will not significantly alter the cluster X-ray emission. There appears to be no escaping the 'over-rich cluster' problem in the standard neutrino model.

We have performed a combined N -body and hydrodynamical simulation for the neutrino scenario using a recently developed particle-based algorithm¹². The method is a combination of the smoothed particle hydrodynamic (SPH) technique^{13,14} with the PP3M N -body code¹⁵. It handles collisional gas and collisionless matter self-consistently, representing each as distinct particle distributions. The particles representing baryon gas elements carry local thermodynamic information along with their mass and kinematic state. Direct interpolation from neighboring particles using a spherically symmetric kernel gives a smoothed estimate of the gas density ρ at each particle's location. Thermal energy ϵ (or equivalently entropy) is treated as an intrinsic particle variable which, combined with the smoothed density estimate ρ in the gas equation of state $P = P(\rho, \epsilon)$ yields the local gas pressure. The pressure gradient acting on a particle is again obtained using kernel estimation of the local pressure field. The smoothing kernel, typically Gaussian, has a finite

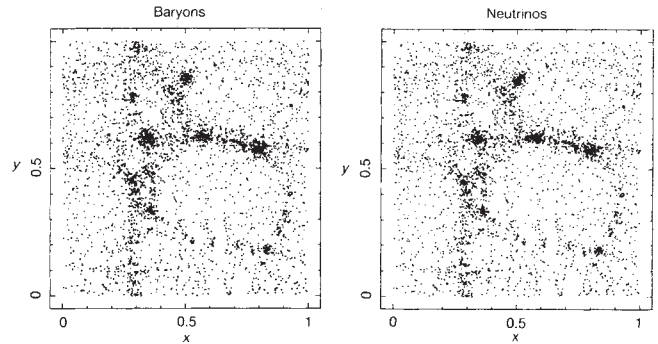


Fig. 1 Projection of the baryon and neutrino particle distributions at a time corresponding to the present epoch.

range which governs the resolution of the calculation. The use of a smoothing scale tied to local conditions provides a wide dynamic range; the three-dimensional hydrodynamics can be followed over several orders of magnitude in density. The code has been tested on problems with semi-analytic solutions such as self-similar expansion and shock-tubes, and has performed well on spherically symmetric collapse problems where results were compared to those from a high resolution one-dimensional lagrangian finite difference code¹². For this problem, we model the baryons as an ideal $\gamma = 5/3$ gas with a constant mean molecular weight, taken to be $\mu = 0.6$, as for a fully ionized primordial hydrogen-helium plasma. Shock heating is incorporated using artificial viscosity. Radiative cooling is unimportant for the bulk of the mass in this problem because cooling timescales are longer than the Hubble time.

We realize the initial mass distribution by random sampling from the spectrum expected after recombination in a neutrino-dominated model with an initial constant curvature ($|\delta_k|^2 \propto k^n$, $n=1$) spectrum of fluctuations³. We call this model with adiabatic initial fluctuations the 'standard' neutrino picture. The spectrum has a large coherence length, $\lambda_c = 13(\Omega h^2)^{-1} \text{ Mpc}$, with very little power on smaller scales. We chose to simulate a cubic region of side $L = 3\lambda_c$ containing two sets of 16^3 particles, one set to represent the mass distribution of neutrinos, the other to represent the baryons. We assume $\Omega = 1$ and $\Omega_b = 0.1$, implying each gas particle represents $m_b = 4.0 \times 10^{11} h^{-4} M_\odot$ of baryons and each collisionless particle represents $m_\nu = 3.6 \times 10^{12} h^{-4} M_\odot$ of neutrinos. The calculation has a resolution limit of $\sim 200 h^{-2} \text{ kpc}$, small enough to resolve the core of a typical rich cluster. The rms density fluctuation within the simulated volume is initially 22%. The gas starts cool, with temperature $T = 10^6 h^{-2} \text{ K}$, characteristic of a typical galactic halo. Little structure develops before the expansion factor R reaches ~ 3 , during which time the gas adiabatically cools to lower temperatures. During the nonlinear phase of clustering, the gas undergoes shock heating. The question we address is whether early rounds of shock heating in pancakes puts the gas on an adiabat high enough to keep it out of the clusters forming at pancake vertices.

We identify the epoch of the first formation of nonlinear structure, an expansion factor $R_* \approx 2.6$, with the epoch of galaxy formation. This should correspond to a redshift $z_* \approx 3$, for consistency with quasar and galaxy observations at this redshift. This implies an expansion factor $R_t = R_*(1+z_*) = 9.6$ to be identified with the present. Figure 1 shows projections of the baryon and neutrino particle distributions in the model at that time. Several rich clusters and a few smaller associations are present. Large, low density voids fill most of the volume. The distribution of baryons appears very similar to that of the neutrinos. To quantify this impression, the local baryon and neutrino densities for each baryonic particle were measured using the kernel estimation technique employed in SPH. Figure

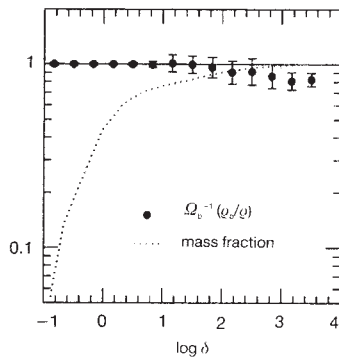


Fig. 2 Local baryon to total mass density, normalized to the global value, as a function of local overdensity. Dotted line is the fraction of baryons with overdensity less than δ .

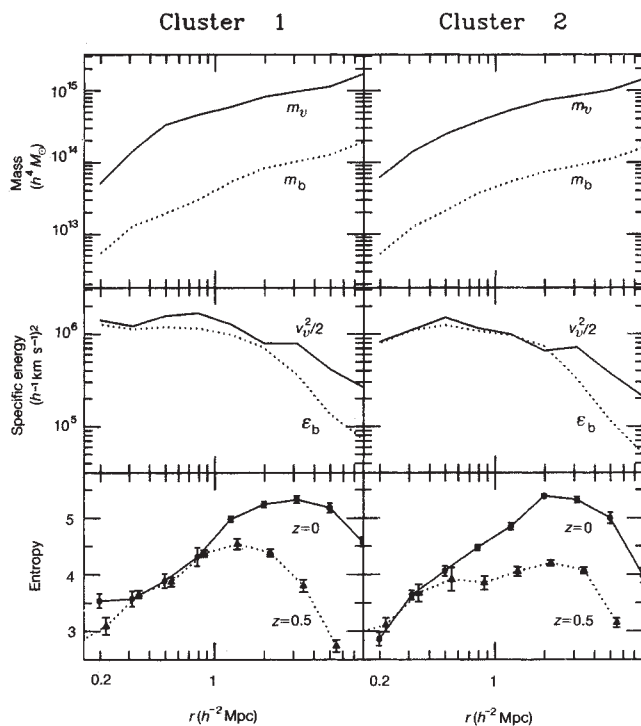


Fig. 3 Cluster profiles for the two largest clusters found in the model portrayed in Fig. 1. The top and middle panels display enclosed mass and specific energy, respectively. Solid lines are neutrino data, dotted lines show the baryon distribution. There is good correspondence between the profiles of baryons and neutrinos in both density and temperature. The lower panels show the entropy profiles of the gas at a redshift $z = 0.5$ (dashed line) and the present epoch (solid line). Means of data binned in logarithmic radial bins are shown with error bars measuring the standard deviations of the data within each bin. The clusters consist of a weakly shocked core with surrounding material shocked onto increasing adiabats.

2 shows the local ratio of the baryon to total mass density, normalized to global value

$$\Omega_b^{-1} \left(\frac{\rho_b}{\rho} \right) \equiv \Omega_b^{-1} \left(\frac{\rho_b}{\rho_b + \rho_\nu} \right) \quad (2)$$

as a function of local overdensity $\delta = \rho/\bar{\rho}$, with $\bar{\rho}$ the mean cosmological density. This quantity should equal unity if no segregation of baryons and neutrinos has taken place. Error bars in the figure denote the standard deviation among particles binned in logarithmic intervals of overdensity. The dotted line in the figure shows the cumulative fraction of baryons as a function of δ . There appears from Fig. 2 to be no significant evidence for large depletions of gas at high overdensities.

Further support for the similarity of the neutrino and baryon distributions comes from examination of individual clusters. Figure 3 shows radial profiles of the enclosed mass and specific energy of both baryons and neutrinos for the two most massive clusters grown in the hydrodynamic model. Also shown are entropy profiles for the gas in the clusters at two epochs, corresponding to the present and to a redshift $z = 0.5$. The most bound particle within each group is the reference point from which profiles are measured. Each cluster contains roughly $10^{15} M_\odot$ within a radius $\sim 3h^{-2}$ Mpc. The baryon mass profiles trace the neutrino profiles well in both clusters, as expected from Fig. 2. The middle panels show the temperature profiles of the neutrinos and baryons. The neutrinos' kinetic energy, $v_\nu^2/2$, where v_ν is the three-dimensional velocity relative to the group centre-of-mass, is compared to the gas thermal energy, $\epsilon_b = 1/(\gamma - 1)kT/\mu m_p$. The gas temperature is in good agreement with the kinetic temperature of the neutrinos, both being roughly isothermal at a temperature equivalent of $T \approx (0.5 - 1) \times 10^8 h^{-2}$ K. This holds over about a decade in radius from the calculation resolution limit of $\sim 200h^{-2}$ kpc to the cluster edge at $\sim 2h^{-2}$ Mpc. Cooler infalling material dominates beyond $2h^{-2}$ Mpc.

The agreement in both density and temperature of the neutrino and baryon distributions within the clusters validates the X-ray luminosity estimates made from purely collisionless models and the assumption that the baryons trace the neutrinos¹¹. Gravitational collapse prior to cluster formation does not generate enough entropy to prevent the bulk of a cluster gas mass from infalling. This is made even clearer from the entropy profiles of the two clusters shown in Fig. 3. Gas in the core is only weakly shocked and is compressed adiabatically as material is added to the cluster. Later infalling material is shocked onto an ever increasing adiabat, generating a strong positive entropy gradient in the cluster to act as a fossil of its shock history.

Feedback from an early generation of radiating objects remains the last hope for the neutrino cluster problem. With adiabatic initial conditions, the first structures to form are near the mass scale of the rich clusters themselves. In this case, feedback can only occur during or after cluster formation. One can argue simply on energetic grounds that this will be difficult. Assume that feedback occurs only after a core of baryonic mass $M_c \approx 10^{13} M_\odot$ collapses and that this core collapses to roughly the density and temperature found in Fig. 3. Lowering the central density by a factor α will require an energy input $E_{\text{inj}} \approx \alpha M_c v_c^2$, where $v_c \approx 1,700h^{-2} \text{ km s}^{-1}$ is the 3-d core velocity dispersion. Even assuming unit absorption of injected energy, this would require an efficient conversion of baryonic rest mass to energy $\epsilon = \alpha v_c^2/c^2 \approx 3 \times 10^{-5} \alpha h^{-2}$. For example, decreasing the density by a factor 5 would require an efficiency $\epsilon \approx 2 \times 10^{-4} h^{-2}$, a value larger than that assumed in explosive models of structure formation¹⁶. The total energy required would be equivalent to about 300 sources radiating 10^{60} erg each, a typical quasar or primeval galaxy energy output. Conventional sources of feedback are therefore unlikely to be effective at significantly lowering the core baryon density. Also, raising the central adiabat too high would result in a convectively unstable cluster which would simply readjust itself to the adiabatic limit after the feedback sources turned off.

Hot dark matter models with initial isothermal perturbations or with added ingredients such as decaying massive particles or cosmic strings are immune to this last argument. Even so, scenarios able to shut down the X-ray production in clusters will have to content with a fine tuning problem which allows just the right amount of X-ray emission from clusters to be consistent with present observations.

We conclude that, for orthodox values of the cosmological parameters H_0 and Ω_b , the standard neutrino picture continues to suffer a fatal 'over-rich cluster' problem, even if methods are invented to coax the galaxy distribution in the model to agree with observations.

This work was supported by the NSF, in part through a NATO Fellowship, and by the SERC. A.E. would like to thank the Department of Astronomy at U.C. Berkeley for hospitality during the course of this work. We also thank S. White for useful critical comments.

Received 5 January; accepted 17 March 1988.

1. Bond, J. R., Efstathiou, G. & Silk, J. *Phys. Rep. Lett.* **45**, 1980-1995 (1980).
2. Doroshkevich, A. G., Khlopov, M. Yu, Sunyaev, R. A., Szalay, A. S. & Zeldovich, Ya. B. *Ann. N.Y. Acad. Sci.* **375**, 32-42 (1980).
3. Bond, J. R. & Szalay, A. S. *Astrophys. J.* **274**, 443-468 (1983).
4. Silk, J., Szalay, A. S. & Zeldovich, Ya. B. *Sci. Amer.* **249**, 72-114 (1983).
5. White, S. D. M., Frenk, C. S. & Davis, M. *Astrophys. J.* **274**, L1-L5 (1983).
6. Centrella, J. & Melott, A. L. *Nature* **305**, 196-198 (1983).
7. Klypin, A. A. & Shandarin, S. F. *Mon. Not. R. astr. Soc.* **204**, 891-907 (1983).
8. Dekel, A. & Aarseth, S. J. *Astrophys. J.* **283**, 1-23 (1984).
9. Shapiro, P. R. in *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M., 447-478 (Reidel, Dordrecht, 1984).
10. Dekel, A. & Rees, M. J. *Nature* **326**, 455-462 (1987).
11. White, S. D. M., Davis, M. & Frenk, C. S. *Mon. Not. R. astr. Soc.* **209**, 27P-31P (1984).
12. Evrard, A. E. submitted to *Mon. Not. R. astr. Soc.*
13. Gingold, R. A. & Monaghan, J. J. *Mon. Not. R. astr. Soc.* **181**, 375-389 (1977).
14. Monaghan, J. J. *Comp. Phys. Rep.* **3**, 71-124 (1985).
15. Efstathiou, G., Davis, M., Frenk, C. & White, S. *Astrophys. J. Suppl. Ser.* **57**, 241-260 (1985).
16. Ostriker, J. P. in *IAU Symp. 130: Evolution of Large Scale Structures in the Universe* (eds Audouze, J. & Szalay, A.) (Reidel, Dordrecht, in the press).
17. Brown, E., Dekel, A. & Shapiro, P. *Astrophys. J.* **328**, 34 (1988).

Particle acceleration by superposition of frequency-controlled laser pulses

Heinrich Hora

Department of Theoretical Physics, University of New South Wales, Kensington, New South Wales 2033, Australia

As the energies of particle accelerators are extended to the TeV range and beyond, the decrease of interaction cross-sections with increasing energy will require that luminosities be increased¹. This will increase the cost of conventional accelerators, and has prompted an examination of alternative designs, incorporating very-high-intensity lasers^{2,3}. One example is the beat-wave accelerator, in which a longitudinal resonance field⁴ produced in a plasma by the beating of two laser beams can accelerate electrons by⁵ $\approx 1 \text{ GeV m}^{-1}$. Unfortunately, this and other types⁶ of laser accelerator involving plasmas encounter problems such as instabilities, self-focusing^{7,8}, de-tuning^{9,10} and the appearance of internal electric fields and double layers^{11,12}, which appear to limit the energy to a few MeV. To overcome these difficulties, I propose here a laser accelerator system that does not require a plasma, in which charged particles are accelerated by nonlinear forces^{13,14} towards minima in a field created by two collinear laser beams, and the laser intensity gradients are moved in phase with the accelerated particles by electro-optical modulation of the frequency and/or phase of the beams. With very high laser powers, electrons from conventional accelerators could be further accelerated by as much as 600 GeV cm^{-1} .

Let us first consider the superposition of two linearly polarized laser waves of different frequencies ω_1 and ω_2 with a phase $\phi(t)$ propagating in the x -direction whose electric fields are described by

$$\mathbf{E}_1 = \mathbf{E}_{10} \cos(\omega_1 x/c - \omega_1 t) \quad (1)$$

$$\mathbf{E}_2 = \mathbf{E}_{20} \cos[\omega_2 x/c - \omega_2 t + \phi] \quad (2)$$

It is no limitation of generality, if the phase for time $t = 0$ is zero:

$$\phi(t=0) = 0; x/c - t = \Phi/\omega_2 \quad (3)$$

The amplitudes perpendicular to the x direction should be equal to the value of $\mathbf{E}_0$. Superposition of the waves (1) and (2) leads to

$$\begin{aligned} \mathbf{E}_1 + \mathbf{E}_2 = 2\mathbf{E}_0 \cos[(\omega_1 + \omega_2)x/2c - (\omega_1 + \omega_2)t/2 - \phi] \\ \cos[(\omega_1 - \omega_2)x/2c - (\omega_1 - \omega_2)t/2 - \phi] \end{aligned} \quad (4)$$

where the first cos-factor nearly corresponds to the initial high-

frequency field of the initial waves, and the second can be considered as an amplitude modification representing a low-frequency wave if the difference between the frequencies is sufficiently small ($\leq 10\%$ in the following examples to be sufficiently close to the maxwellian exact solution).

The aim with respect to the acceleration of charged particles in the field of the two laser waves *in vacuo* is to produce a strong gradient of the $\mathbf{E}$ -field along the x -direction, such that a particle with charge Ze and mass m in this gradient will receive an acceleration in the x -direction of the ponderomotive force approximation of the general nonlinear force¹³:

$$\mathbf{F} = -\nabla E^2 / (16\pi n_{ec}) \quad (5)$$

where n_{ec} is the cutoff density of the particles with respect to the average frequency of the two superposed laser beams. This gradient should move with the (temporally and spatially changing) speed of the charged particle to be accelerated.

If the frequencies and the phase in equation (4) (apart from the transient processes of switching on and off¹⁴ represented by wave packets) are constant, the gradient (5) moves with the vacuum speed of light. To achieve a lower (or in the initial steps very much lower) speed $v(t)$, the time dependence of the frequency ω_2 , or of both frequencies (or by a temporal variation of the phase ϕ being not discussed here) can be used according to the speed of the charged particles. The fit of the velocity v with the motion of the field gradient determining the nonlinear force is given by

$$v(t) = [(\omega_1 - \omega_2)/2 - t(\partial\omega_2/\partial t)/2 + 0.5(\partial\phi/\partial t)]2c/(\omega_1 - \omega_2) \quad (6)$$

One sees that v reaches the speed of light if the frequency and the phase do not depend on time, whereas the velocity zero requires, as the other extreme case, that the square bracket term be zero.

With respect to the achievement of the desired velocity-controlled propagation of the long wavelength field gradient needed for the particle acceleration, a slow time-dependence of the frequency and/or the phase is taken into consideration. The solutions (2) and (3) are then no longer solutions of the wave equation for the vacuum, or of the maxwellian equations. By re-substituting and spatial and temporal evaluation one finds the limitation under which the time-dependence of the frequency and/or the phase is a good approximation. The result is that

$$\begin{aligned} |\omega_2| \gg \left| \frac{\partial\omega_2}{\partial t} \frac{x}{c} - \frac{\partial\omega_2}{\partial t} t + \frac{\partial\phi}{\partial t} \right| \\ = \left| \frac{\partial\omega_2}{\partial t} \left(-\frac{\Phi(t)}{\omega_2(t)} + \frac{\Phi(0)}{\omega_2(0)} \right) + \frac{\partial\phi}{\partial t} \right| \end{aligned} \quad (7)$$

and

$$\begin{aligned} |\omega_2|^2 \gg \left| -2 \frac{\partial\omega_2}{\partial t} + \frac{\partial^2\omega_2}{\partial t^2} \frac{x}{c} - \frac{\partial^2\omega_2}{\partial t^2} t + \frac{\partial^2\phi}{\partial t^2} \right| \\ = \left| \frac{\partial^2\omega_2}{\partial t^2} \left(-\frac{\Phi(t)}{\omega_2(t)} + \frac{\Phi(0)}{\omega_2(0)} \right) - 2 \frac{\partial\omega_2}{\partial t} + \frac{\partial^2\phi}{\partial t^2} \right| \end{aligned} \quad (8)$$

has to be fulfilled. In view of the necessary smallness of the right-hand side of relation (7) because of equation (6) for any positive velocities below c , the condition (7) is easily fulfilled. The same is the case for relation (8) if non-pathological second derivatives of the functions involved are considered.

From equation (6), the same time variation of the frequency and/or the phase change is relatively slow, well within the means of today's electro-optical processes that are fully controlled by the necessary phasing of the optimum acceleration of the particles with the laser fields. Figure 1 shows an arrangement to produce the timescale frequency and/or phase shift of the second frequency in a controlled way according to equation (6). The first laser beam from the laser L_1 should be a vertical beam passing a 45° mirror (M), which for the frequency of the first laser should ideally be transparent. The light from the second

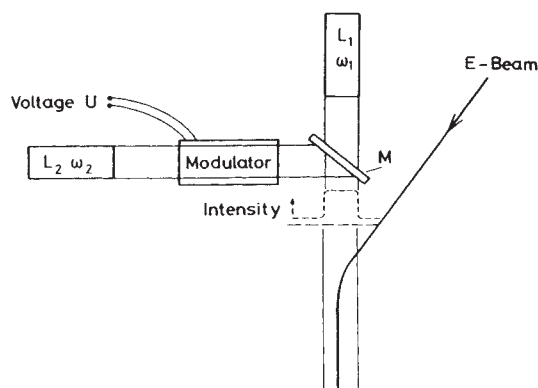


Fig. 1 Superposition of laser beams L_1 and L_2 with frequencies ω_1 and ω_2 respectively by a mirror M to produce an accelerator field for electrons (E-Beams) to be injected. The voltage-controlled electro-optical modulator varies the frequency ω_2 for the acceleration process.

tunable laser L_2 in a horizontal direction may have frequencies within a range for which M is close to 100% reflection for radiation incident at 45° , joining the beams collinearly to fulfil the conditions of equations (1) and (2) in the space below M , where the same laser intensities of the beams are assumed. An electro-optical modulator is placed between the second laser L_2 and the mirror whose refractive index can be changed, for example, by a transversal electric field due to the Kerr effect from varying a voltage U at the plates at the upper and lower sides of the modulator to temporally vary the phase. The time-dependence of the phase in equation (6) can then be controlled very precisely by the time-dependence of U , especially if the modulator is appropriately long. For the time-dependence of the tuning of the frequency, similar experimental conditions are given.

If a (pulsed or bunched) particle beam, such as an electron beam, is then fitted into the laser field of equations (4) to (6), as shown in Fig. 1, by a magnetic field turning the direction of the particles into the collinear geometry of the laser beams, and if the right phasing between the bunches and the gradients of the laser field are optimized, the particles will be accelerated.

The problems of this method will be apparent when the limitations of the approximation presented are violated. One has to consider the finite diameter of the beams (instead of the infinitely spread plane waves assumed in the preceding equations) represented by an intensity profile, as drawn by the dotted lines in Fig. 1, for example, or an axial intensity minimum (hollow beam). Even if there is a nearly plane wave approximation in the centre of the beam, the decay at the border areas will produce deviations on the laser field, even with longitudinal components of electric E and magnetic H field vectors of the laser field¹³. Further discussion involves the transient processes of the laser pulses¹⁴.

Furthermore, if the particle density is high, the dielectric response of the particles to the laser beam will cause the more general force densities known from the nonlinear force^{13,14}:

$$\mathbf{f} = \mathbf{v} \times \mathbf{H}/c + \mathbf{E} \nabla \cdot \mathbf{E}/4\pi + [1 + (\partial/\partial t)/\omega] \nabla(\tilde{n}^2 - 1) \mathbf{E} \mathbf{E}/4\pi \quad (9)$$

Further, the space charge effects of the particle bunches have to be taken into account as when considering the behaviour in plasmas^{11,12}, for which extensive numerical codes have been developed to simulate the real conditions, and the relativistic description has to be included. For very short laser pulses or short particle bunches, a nonlinear energy transfer to the particles will occur to stretch the bunches¹⁵. The switching of processes on and off, as described in equation (9) by the bracket with

the time derivative, is essential for a very large number of particles and for the very-high-energy transfer for high-energy accelerators. To counteract the divergence of the particle beams, the laser intensity profile (dashed in Fig. 1) could be chosen to have a minimum at the axis, such that a nearly hollow beam is produced focusing the particles by nonlinear forces directed to the beam centre.

The net energy gain G per length for electrons by the nonlinear forces in the optimized points of field gradients is approximated as

$$G = 1.758 \times 10^{-4} P/(\lambda AB^2) \text{ eV cm}^{-1} \quad (10)$$

where P is the laser power of both collinear beams in watts, λ is the laser wavelength in cm, A is the ratio of wavelengths of the second cos function to the first one in equation (4), and B is the number of wavelengths (of the wave of equation (1)) of the beam diameter in the interaction area for the electrons. Taking laser pulses of 250 nm wavelength, and an extreme, but feasible, case of 100 TW laser pulse power and $A = B = 10$, an acceleration of 600 GeV cm^{-1} will result.

Received 8 January; accepted 31 March 1988.

1. Hora, H. *Atomkernenergie Kerntechnik* **48**, 185-189 (1986).
2. *Laser Accelerators AIP Proc. No. 130* (Joshi, C. & Katsouleas, T.) (Am. Inst. Phys., New York, 1985).
3. *New Developments in Particle Acceleration Techniques Proc. CERN 87-11 ECFA-87/110 Vols 1 and 2* (ed. Turner, S.) (CERN, Geneva, 1987).
4. Tajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267-270 (1979).
5. Tajima, T. *Laser and Particle Beams* **3**, 351-413 (1985).
6. Riege, H. *New Developments in Particle Acceleration Techniques Proc. CERN 87-11 ECFA-87/110 Vol. 2*, 554-572 (CERN, Geneva, 1987).
7. Hora, H. *Z. Phys.* **226**, 156-159 (1969).
8. Hora, H. *J. Opt. Soc. Am.* **65**, 882-886 (1975).
9. Salomaa, R. R. E. & Karttunen, S. J. *New Developments in Particle Acceleration Techniques Proc. CERN 87-11 ECFA-87/110 Vol. 2*, 458-471 (CERN, Geneva, 1987).
10. Pesme, D. *et al. Laser and Particle Beams* **6**, No. 2 (1988).
11. Eliezer, S. & Ludmirski, A. *Laser and Particle Beams* **1**, 251-270 (1983).
12. Hora, H., Lalouis, P. & Eliezer, S. *Phys. Rev. Lett.* **53**, 1650-1652 (1985).
13. Hora, H. *Phys. Fluids* **12**, 182-191 (1969); *Physics of Laser Driven Plasmas* (Wiley, New York, 1981).
14. Hora, H. *Phys. Fluids* **28**, 2705-2706 (1985).
15. Batchelor, M. T. & Stening, R. J. *Laser Particle Beams* **3**, 189-196 (1985).

Geomagnetic activity during the passage of the Earth through Halley's tail in 1910

C. T. Russell*, M. von Dornum*, R. L. McPherron*, J. H. Allen† & L. Morris†

* Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

† National Geophysical Data Center, NOAA/NESDIS, Boulder, Colorado 80303, USA

In 1910 Halley's comet as seen from the Earth transited the Sun and hence the Earth should have passed through the comet's tail. From our understanding of the nature of a cometary tail we would expect the tail to have shielded the Earth's magnetic field from the dynamic pressure of the solar wind except for a brief period in the centre of the tail when the Earth passed through the dense flowing ion tail. Such a disturbance occurred, but it appears to have been 12 h too early. A detailed study of the records reveals that the discrepancy is due to a change in the convention for determining the start of the day. In 1910 the commonly used Greenwich Mean Time referred to Astronomical Time at Greenwich, which was 12 h different from the Greenwich Mean Time, or Universal Time which is in use today. Hence the disturbance associated with Halley did in fact arrive at the expected time and no unusual aberration of the solar wind need be invoked to explain the timing. The disturbance consisted of two troughs in the horizontal component of the Earth's magnetic field separated by ~ 14 h

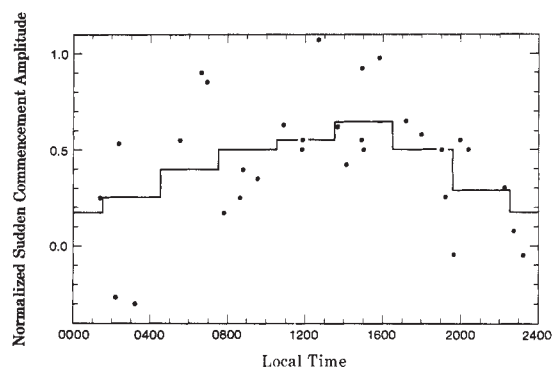


Fig. 1 The amplitude of sudden commencements in 1910 normalized by the published amplitudes⁵ and plotted versus local time. In constructing this plot we assumed that the original records and tables were given in Greenwich Mean Astronomical Time which is 12 h different from the modern definition of Greenwich Mean Time. The histogram shows the median amplitude for 6-h intervals overlapped by 3 h centred on the times shown.

presumably associated with wakes in the solar wind momentum flux on either side of the ion tail. The disturbance was independent of latitude indicating that the responsible current system flowed far above the surface of the Earth. After the passage of the comet, the magnetosphere was left in a mildly disturbed condition with a weak ring current present.

On 19 May 1910 at 04:00 UT comet Halley transited the Sun as seen from Earth. Comet Halley was 0.16 AU closer to the Sun at this time. If the solar wind were flowing at 500 km s^{-1} radially out from the Sun the Earth should have passed through Halley's tail at 12:00 UT. Various studies indicate a geomagnetic disturbance at about this time. Within a week of Halley's passage in 1910 Chree¹ had published an account of a weak magnetic disturbance in association with the crossing of Halley's tail. Many years later Ivanov and Shevnev² examined data from seven Northern Hemisphere stations and concluded that Halley had a geomagnetic effect. More recently, Yan and co-workers³ found disturbances in the Lu-Kia-Pang observatory records. Converting to modern Universal Time, as we will discuss below, these disturbances occurred as three troughs in the horizontal magnetic field at 02:00 UT, at 15:00 UT and a shallow one at midnight on 19–20 May. Finally, Russell and co-workers examined the aa-index and the horizontal component from four stations: Agincourt, Cheltenham, Kew and Falmouth⁴. The aa-index showed that the disturbance was not obviously recurrent at the rotation rate of the Sun and thus possibly not solar related.

In the Russell *et al.* study the four sub-auroral stations showed a disturbance that consisted principally of two troughs centred about 12 h before the expected crossing of the ion tail. These troughs were interpreted to be due to the shielding of the Earth's magnetosphere by the wake of the comet on either side of the ion tail but an unusual direction of the solar wind had to be invoked to explain this early arrival. To understand the unusual nature of the observed variation we sought to obtain more data. These additional data from Tbilisi, Sverdlovsk, Ebro, Batavia, Lu-Kia-Pang, Apia, Honolulu, Tucson and Vieques enabled us not only to understand the cometary interaction better but also revealed a problem of misunderstanding of the timing of all pre-1925 magnetograms.

The initial data we obtained were photocopies of the original station magnetic tracings. These data (from Agincourt, Cheltenham, Kew, Falmouth, Tbilisi and Sverdlovsk) were manually digitized and then analysed by removing the 'quiet-day' variation. These analyses showed that the major variations of the magnetic field were slow enough that they could be followed by using the hourly values recorded in station logs; thus, we analysed the data from Ebro, Batavia, Lu-Kai-Pang, Apia, Honolulu, Tucson and Vieques. There were some inconsistencies

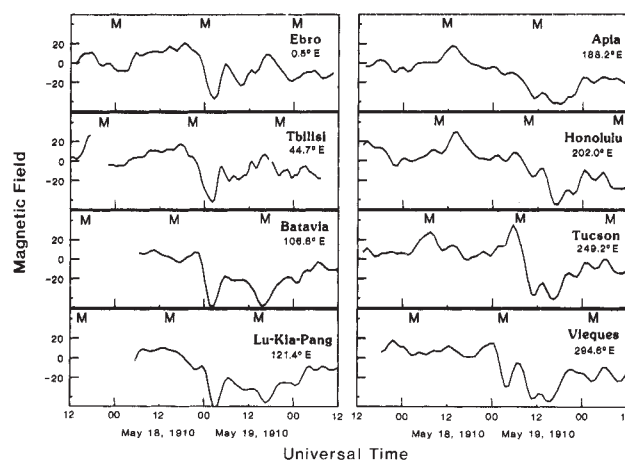


Fig. 2 The deviation from quiet-day values of the horizontal component at eight mid-latitude magnetic observatories during the 1910 passage of the Earth through Halley's tail. The 'correction' to modern GMT or Universal Time has been made in these data. Local midnight at each station is denoted by the letter 'M'. The observations for Tbilisi have been plotted with our best estimate of the scale factor for this station.

in these data so we compared the onset times of sudden commencements in 1910 as recorded by Mayaud⁵ with the times observed in the station logs. These comparisons generally revealed good agreement if all the stations except Apia recorded instantaneous values at a time slightly after the hour while Apia recorded an hourly average. One station, Vieques, differed by 4 h. Its data had been recorded in local time. This analysis, *per se*, brought us no closer to resolving the 12-h discrepancy.

As noted in the Russell *et al.* study Astronomical Time had changed by 12 h in 1925⁴; this was also recently noted in the 1910 Halley Atlas⁶, but no-one with whom we consulted would admit to such a time shift for geomagnetic data.

It is possible to use the properties of the 1910 geomagnetic records to check what time system was in use because geomagnetic variations at midnight local time differ significantly from those at noon. A slight addition to our sudden-commencement analysis provided us with such a test because sudden-commencement amplitudes are greatest at noon and least at night. Thus we analysed the amplitudes of all major sudden compressions in 1910 as seen at our stations. These were then normalized by the amplitude in Mayaud's table⁵, and are plotted in Fig. 1 assuming there was a 12-h shift in GMT between 1910 and now. The reason that the maximum amplitudes tend to be less than unity is that we did not have the original records and thus could examine only the hourly values from the station logs rather than

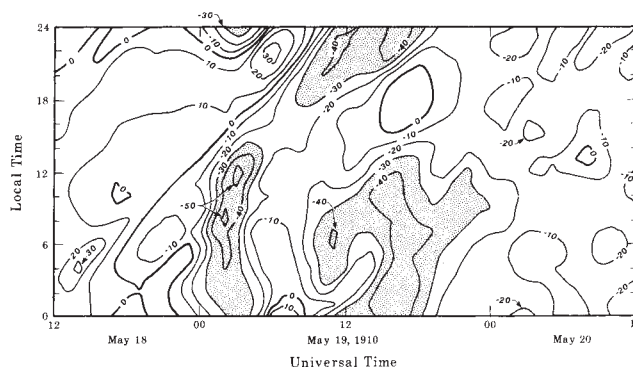


Fig. 3 Deviation of horizontal component of magnetic field from quiet day. The observations presented as time series in Fig. 2 displayed as a contour map in Local Time-Universal Time coordinates. The isocontours every 10 nT are shown.

the instantaneous values of the jumps. The amplitudes shown in Fig. 1 maximize near local noon confirming our suspicion that in 1910 the expression Greenwich Mean Time refers to the Astronomical Time at Greenwich which was then 12 h different from Universal Time now and what we call Greenwich Mean Time today. Not only does this observation have important implications for our analysis of the 1910 Halley data but it also presumably affects all geomagnetic records before 1925 and in particular the aa-index which is supposedly a homogeneous time series going back to 1868. The 12-h change was not noted by Mayaud⁵ but clearly is present because the times of the sudden commencements in his lists agree with the 1910 definition of GMT.

Returning to our analysis of comet Halley's tail passage we can plot the deviation of the horizontal component of the magnetic field from the quiet-day values for all our mid-latitude stations. We do not include the data from Sverdlovsk here because it is a sub-auroral observation but its variations were similar to those shown at Tbilisi. This is shown in Fig. 2 with corrected times. The principal worldwide variation is the occurrence of two troughs as noted by Yan and co-workers³. There are, however, important deviations from this behaviour for stations in the dusk sector. This can be seen more clearly in Fig. 3 where we contour the variations in the Local Time–Universal Time plane. The deepest parts of the magnetic troughs have been shaded. We see that in the late afternoon and early evening hours the magnetic field is not as depressed as from midnight through the early afternoon hours. We attribute this to the collapse of the magnetotail as often occurs during substorms⁷. We further note that the sizes of these disturbances are similar to those found in sub-auroral records⁴, attesting to the global nature of the disturbance as would occur if the responsible currents flowed on the Earth's magnetopause rather than in the auroral ionosphere.

Another way of examining such variations is to create a Dst or ring current index by finding the average horizontal field deviation around the world at mid-latitudes. We can do this from the data in Fig. 3 by giving all local times equal weight. This 1910 Dst index is shown in Fig. 4. It shows the two troughs

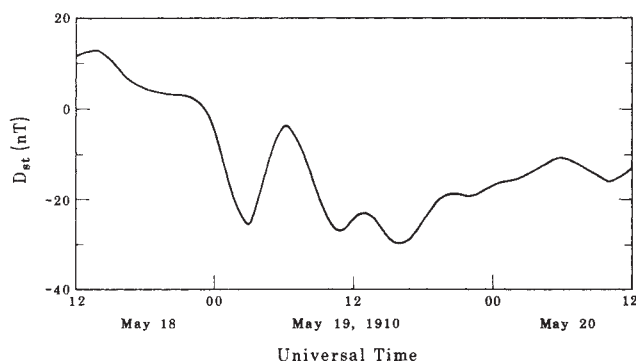


Fig. 4 The observations displayed in Fig. 3 averaged over local time to give a Dst index.

as seen in the original records but also shows a weak depression in the field persisting after the time of the second trough. This could be due to some ring current energization associated with the cometary passage, some prolonged deficit in the solar-wind momentum flux caused by the comet or the material left behind by the comet in its orbit, or possible just due to normal solar wind effects.

In short, careful analysis of the 1910 geomagnetic records indicates that GMT as used by observatories had a different definition than now. From this knowledge we find that the disturbances seen on the Earth on 19 May 1910 were very probably associated with the entry of the Earth into the momentum flux wake of the comet on either side of a flowing ion tail.

The timing difference found affects all the records examined in 1910 including the aa-index and presumably affects all geomagnetic records until 1925. Researchers should be aware of this.

This work was supported by the National Aeronautics and Space Administration.

Received 2 February; accepted 17 March 1988.

1. Chree, C. *Nature* **83**, 367–368 (1910).
2. Ivanov, K. G. & Shevchin, A. D. *Geomag. Aeron.* **5**, 634–637 (1966).
3. Yan, L.-S., Li, Z.-Y. & Gu, S.-Y. *Chin. Astron. Astrophys.* **8**, 260–264 (1984).
4. Russell, C. T. et al. *J. geophys. Res.* **92**, 11195–11200 (1987).
5. Mayaud, P. N. *A Hundred Year Series of Geomagnetic Data 1868–1967* (IUGG Publications Office, Paris, 1973).
6. Donn, B., Rahe, J. & Brandt, J. C. *Atlas of Comet Halley 1910/11*, NASA SP-488 (National Aeronautics and Space Administration, Washington, DC, 1986).
7. Russell, C. T. & McPherron, R. L. *Space Sci. Rev.* **15**, 205–266 (1973).

Effects of sulphur deposition on lake-water chemistry in Ontario, Canada

B. P. Neary & P. J. Dillon

Dorset Research Centre, Ontario Ministry of the Environment,
PO Box 39, Dorset, Ontario P0A 1E0, Canada

Clear links between regional lake acidification and the deposition of strong acids from the atmosphere have not been reported for North America^{1,2}. This is unlike the situation in Scandinavia^{3–5}, where acidification of lakes on a regional basis corresponds closely with the sulphur deposition pattern. Sulphur deposition in Ontario, Canada, decreases from south to north⁶, with localized high deposition rates found in the Sudbury region^{7,8}. We compared the chemical attributes of 1,168 lakes in different sulphur deposition zones defined by increments of $0.25 \text{ g S m}^{-2} \text{ yr}^{-1}$. As sulphur deposition increased, lake SO_4^{2-} concentration increased, and alkalinity concentration and pH decreased. Increasing importance of organic acids did not result in lower alkalinity or pH. In the zones with sulphur deposition $>1.0 \text{ g m}^{-2} \text{ yr}^{-1}$, almost one quarter of the lakes had pH values that would be expected to result in adverse biological effects.

To assess the effects of sulphur deposition on lake chemistry on a regional scale, we first divided Ontario into zones based on total sulphur-deposition fields (Fig. 1a) reported for 1983 (ref. 6). This deposition map was superimposed on a map of tertiary watershed boundaries, and six groupings of watersheds were constructed to match as closely as possible the sulphur deposition zones (Fig. 1b). A seventh zone was constructed around Sudbury. This zone, $200 \text{ km} \times 200 \text{ km}$, includes all those lakes believed to have been affected by the historically very high sulphur emissions from the local smelters^{9,10}. Because of uncertainty in the magnitude of these emissions and subsequent deposition rates¹¹ and because sulphur deposition in this area has changed very substantially within the past decade¹², the deposition in zone 7 was not quantified, but was considered to be the highest level in Ontario. In addition, the effects of both historical and current sulphur emissions on lake-water chemistry in this zone cannot be adequately separated from the influence of long-range sulphur transport. Unlike zone 7, the sulphur deposited in the other zones is attributable to long-range transport, principally from southern Ontario and the northern United States^{13,14}.

Chemical data included alkalinity (modified Gran titration), pH (combination glass electrode), SO_4^{2-} (ion chromatography), base cations (atomic absorption), dissolved organic carbon (DOC; ultraviolet digestion and colorimetry), and conductivity. These were available for a total of 1,855 lakes (after eliminating those where ion charge, including organic anion concentration calculated from DOC and pH according to Oliver¹⁵, was unbal-

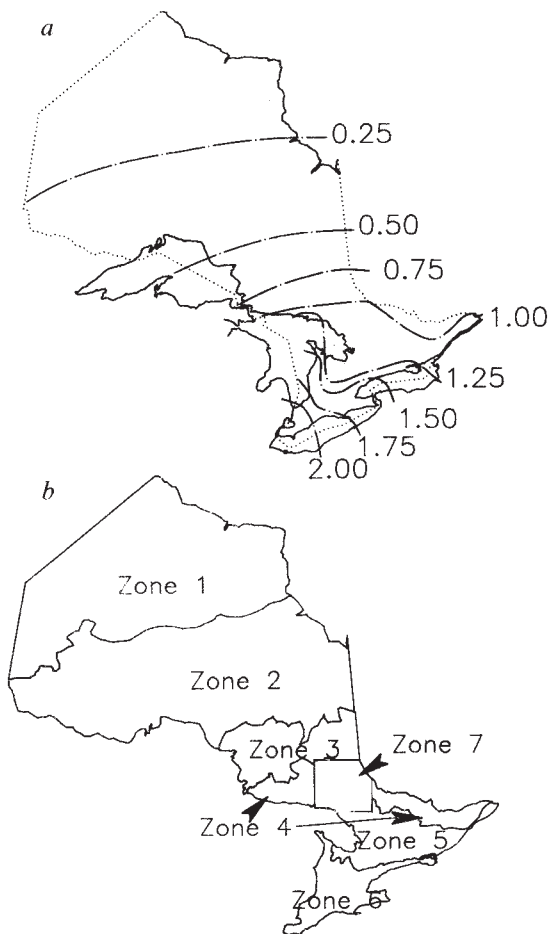


Fig. 1 *a*, Total sulphur deposition ($\text{g S m}^{-2} \text{yr}^{-1}$) measured in Ontario in 1983⁶. *b*, The division of Ontario into deposition zones by overlaying of major watershed boundaries with deposition fields. A separate zone (200 km $\times$ 200 km) was created with Sudbury at the centre because of the well-documented local effects. There are very few lakes in zone 6; all are hard water.

anced by >10%) throughout Ontario. These lakes, which were surveyed between 1981 and 1987, were categorized by sulphur deposition zone. There were a significant number of lakes in each deposition zone except for zone 6; this region of Ontario has very few lakes in total (all of which are hard-water lakes) and was excluded from subsequent analysis.

The data set was further reduced by considering only the soft-water lakes, defined on the basis of a bimodal conductivity distribution as those lakes with specific ion conductivity $\leq 50 \mu\text{S}$. In doing this, we are, of course, biasing the data set to include those that could be considered sensitive to acidic deposition, because the lakes with conductivity $> 50 \mu\text{S}$ are, with very few exceptions (for example, from road-salt effects), those with appreciable carbonate substrate in their catchments. However, our purpose here is to quantify the relations between sulphur deposition and water chemistry of sensitive lakes rather than to assess the scope of the effects of acid deposition. As a result of this selection criterion, the data set was reduced to 1,168 lakes. This subset constituted almost two thirds of the original data set.

The chemical characteristics of the soft-water lakes in the six deposition zones are summarized in Table 1. The trend to increasing contribution of SO_4^{2-} and declining importance of bicarbonate (and therefore alkalinity) as sulphur deposition increases is clearly evident; SO_4^{2-} increased from ~11% of the anion content in zone 1 to 64% in zone 5 and 80% in the Sudbury region (zone 7). The organic anion contribution (A^-)

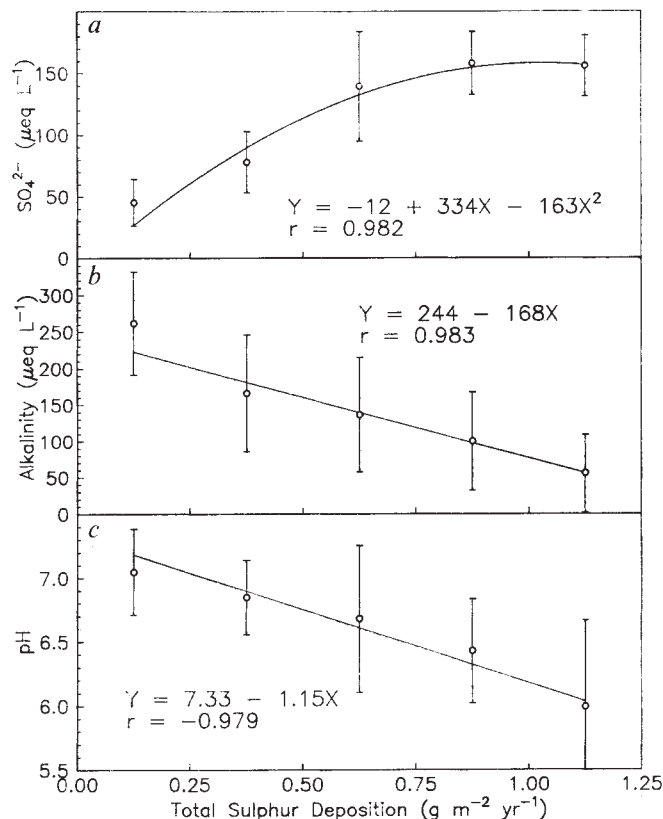


Fig. 2 Relation between sulphur deposition and *a*, mean SO_4^{2-} concentration; *b*, alkalinity concentration and *c*, pH of lakes in zones 1-5. Zone-7 lakes had the highest SO_4^{2-} and lowest alkalinity and pH (that is followed the same trends), but are not shown on the figure because the sulphur deposition in this zone is uncertain. The regression lines are weighted by the inverse of the standard error; the mid-points of the deposition ranges, for example, $0.125 \text{ g S m}^{-2} \text{yr}^{-1}$, were used as the x-variable. Error bars correspond to 1 standard deviation; $P < 0.002$ in all cases.

ranged only from 12 to 24%, with the highest proportion occurring in the two zones of lowest sulphur deposition. The increase in SO_4^{2-} importance as sulphur deposition increased was balanced by a decrease in HCO_3^- , which dropped from 62% in zone 1 progressively to 20% in zone 5 and 10% in zone 7. The observation that, in zones where sulphur deposition was low, the lakes with the highest A^- did not have low pH or alkalinity (for example one highly coloured lake in zone 1 had pH of 5.60 and alkalinity of $72 \mu\text{eq l}^{-1}$; this was the only lake of 169 in zones 1 and 2 with pH < 6.0), indicates that organic acids are not major contributors to the observed trends in pH and alkalinity.

The relation between sulphur deposition and the mean SO_4^{2-} concentration of lakes in each deposition zone is shown in Fig. 2*a*. Although there is a strong correlation between sulphur deposition and SO_4^{2-} concentration (2), zones 4 and 5 (sulphur deposition between 0.75 and $1.25 \text{ g m}^{-2} \text{yr}^{-1}$) had almost identical SO_4^{2-} concentrations. This is perhaps because SO_4^{2-} reduction is stimulated at high sulphur input rates^{16,17}; in fact, SO_4^{2-} reduction is a significant source of alkalinity in lakes receiving high deposition. The Sudbury zone (7), with the highest sulphur deposition had the highest SO_4^{2-} concentration, but was not included in the regression analysis because the actual total sulphur deposition is not well known. By weighted-polynomial least-squares regression, an estimate of $< 0 \mu\text{eq l}^{-1}$ for the SO_4^{2-} concentration of lakes where there is no sulphur deposition was obtained. This estimate of the lower limit of the 'background' SO_4^{2-} concentration in soft-water lakes, although made using

the unrealistic assumption that 'background' sulphur deposition is zero does imply that geological sources of SO_4^{2-} are inconsequential for the soft-water lakes. Either the mean measured value ($45 \mu\text{eq l}^{-1}$) of lakes in the lowest deposition zone or the minimum measured values (several lakes had SO_4^{2-} of $\sim 17 \mu\text{eq l}^{-1}$) could be considered as estimates of 'background' SO_4^{2-} . These results are consistent with estimates made by other techniques¹⁸.

The relations of sulphur deposition with pH and with alkalinity are shown in Fig. 3b, c. The very strong negative correlation of sulphur deposition with both suggests that increasing sulphur deposition has resulted in a reduction in alkalinity and pH in lakes. Such relations were not reported in the analysis of other surveys^{1,2}, perhaps because sensitive (low conductivity) lakes were not separated. Again, the zone-7 lakes follow the same pattern with lowest pH and alkalinity. In zone 5, the average pH is slightly less than 6.0, and is close to the pH where biological effects are expected^{19,20}. Of the 557 lakes surveyed in this zone, 23% had $\text{pH} < 5.7$, and 22% had alkalinity $< 20 \mu\text{eq l}^{-1}$, levels indicative of significant biological damage²¹.

A comparison of the alkalinity distributions found in zones 1 and 5 indicated that a shift in the alkalinity median from $260 \mu\text{eq l}^{-1}$ in zone 1 to $43 \mu\text{eq l}^{-1}$ in zone 5 had occurred. Although this is a measure of the effects of the increased sulphur deposition in the latter zone, other factors are involved.

Although all lakes considered in this analysis were soft water, the average base cation content of lakes in zone 1 was greater than that in other zones. To quantify the importance of sulphur deposition in the loss of alkalinity it is necessary to account for the higher base-cation content in zone 1. For example, the decline in average alkalinity from zone 1 to zone 5 ($208 \mu\text{eq l}^{-1}$) is partly caused by a decline in base cations ($138 \mu\text{eq l}^{-1}$) which is, in turn, partly balanced by a reduction in organic anions ($55 \mu\text{eq l}^{-1}$). The net effect is equivalent to a loss of $138 - 55 = 83 \mu\text{eq l}^{-1}$ of alkalinity, 40% of the total decline, with the remainder ($\sim 60\%$) balanced by the rise in SO_4^{2-} ($111 \mu\text{eq l}^{-1}$). Even the soft-water lakes of zone 2, where sulphur deposition was $0.25\text{--}0.50 \text{ g m}^{-2} \text{ yr}^{-1}$, have lower alkalinity and pH than zone 1. The data indicate that all lakes in zones 2–5 have lost a significant portion of their alkalinity relative to those in zone 1. The decreasing pH and alkalinity that coincides with increasing sulphur deposition cannot be attributed to increasing levels of organic acids. In fact, the opposite trend is noted (Table 1); the highest organic anion levels are found in zone 1, the lowest in zones 4, 5 and 7.

Table 1 Mean and range of concentrations of major ions ($\mu\text{eq l}^{-1}$), pH , conductivity (μS), and organic anions (A^- ; $\mu\text{eq l}^{-1}$) in soft-water lakes in six deposition zones in Ontario

	Deposition zone					
	1	2	3	4	5	7
<i>n</i>	37	132	72	268	557	102
<i>pH</i>	7.05	6.85	6.68	6.43	5.99	5.79
	5.60–7.53	6.09–7.46	4.79–7.44	4.85–7.45	4.22–7.01	4.09–7.06
Cond.	39.3	33.2	36.9	34.9	31.6	37.7
	20.0–49.0	16.0–50.0	17.7–49.8	18.8–50.0	17.0–50.0	20.0–49.0
A^-	95	75.6	53.5	43.5	40.2	31.7
	38.9–165	20.6–203	4.5–158	8.4–137	10.9–126	3.3–66.3
SO_4^{2-}	45.3	78.1	139	158	156	220
	16.9–90.5	16.4–184	68.9–260	87.4–229	40.8–231	97.0–364
Alk.	262	166	136	100	54.4	24.4
	72–390	38.4–429	–15.8–321	–6.4–308	–65.8–294	–98.8–135
Ca^{2+}	224	185	215	157	140	147
	115–324	49.9–399	99.8–399	61.9–339	49.9–369	69.9–240
Mg^{2+}	95.2	83.2	73.7	82.6	60.0	73.6
	41.9–148	44.4–165	26.3–132	33.7–156	23.0–153	31.3–123
Na^+	48.0	36.8	30.4	36.7	32.4	37.8
	15.4–79.1	9.2–74.7	11.4–61.5	14.5–70.3	13.2–81.3	17.6–83.5
K^+	15.5	11.4	8.9	13.5	11.4	11.6
	5.4–28.1	1.3–25.6	3.1–15.3	3.6–41.9	2.6–29.7	5.1–26.9

There are few lakes in zone 6, and none are soft water; $n=1,168$. Cond., conductivity. Alk., alkalinity.

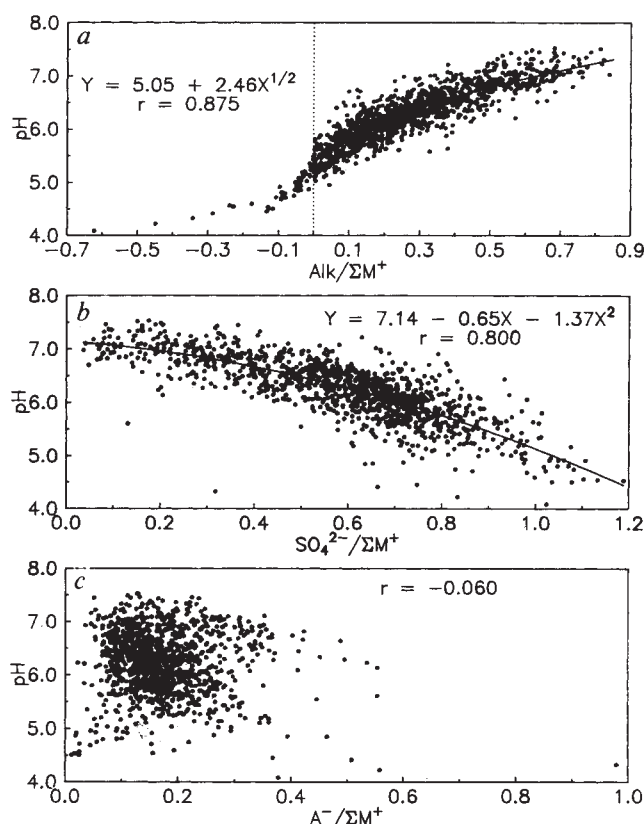


Fig. 3 Relation between pH and the ratio of *a*, alkalinity; *b*, SO_4^{2-} and *c*, organic anions (A^-) with base cations $\text{M}^+ = (\text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+)$. Minor ($< 1\%$) contributors to the cations (for example NH_4^+) were excluded ($n=1,168$ lakes). The empirical $\text{alk}/\Sigma\text{M}^+ - \text{pH}$ relation is based on only those lakes with alkalinity > 0 and $\text{pH} > 5.0$ ($n=1,129$); at lower alkalinity, pH controls the alkalinity and autocorrelation results.

The ratio of each of SO_4^{2-} , organic anions and alkalinity to base cations ($\text{M}^+ = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+$) is plotted against pH in Fig. 3. As $\text{Alk}/\Sigma\text{M}^+$ increases, pH increases as expected. (At negative alkalinities, the relation is spurious because H^+ dominates the alkalinity term.) But the very strong relation between $\text{SO}_4^{2-}/\text{M}^+$ and pH is expected only if increases in lake- SO_4^{2-} concentrations coincide with decreases in pH . That organic acids do not, in general, control lake pH is evident from the lack of any relation between $\text{A}^-/\Sigma\text{M}^+$ and pH . Only 1 lake out of 1,168 had $\text{pH} < 6.0$, low SO_4^{2-} and high A^- .

By selecting only soft-water lakes for this analysis (1,168 out of a total of 1,855 lakes surveyed) much of the variability in lake chemistry attributable to geological effects has been removed. The results presented here demonstrate that the pattern of increasing sulphur deposition from north to south in Ontario results in an analogous pattern of increasing lake SO_4^{2-} concentration, but most important, a pattern of decreasing lake alkalinity and pH . Clear links between sulphur-deposition patterns and lake pH and alkalinity patterns have not been observed in North America previously. The presence of organic acids does not explain the observed pH and alkalinity patterns; in fact, pH and alkalinity are highest where organic anion levels are highest. If it is assumed that the chemical characteristics of the lakes in the lowest deposition zone ($< 0.25 \text{ g S m}^{-2} \text{ yr}^{-1}$) are indicative of 'background' conditions, then substantial alkalinity has been lost in all other regions. A quarter of the sensitive lakes surveyed where sulphur deposition was between 1.0 and $1.25 \text{ g m}^{-2} \text{ yr}^{-1}$ had $\text{pH} < 5.7$ and alkalinity $< 20 \mu\text{eq l}^{-1}$, chemical conditions consistent with biological damage.

This work was carried out as part of the Acidic Precipitation in Ontario Study. We thank D. Dodge and D. Wales for their

assistance in obtaining portions of the data base, Bill Keller, Len Maki and Ontario Ministry of Natural Resources personnel for sample collection, B. LaZerte for critical review of the manuscript, and E. Gorham for very helpful suggestions.

Received 15 October 1987; accepted 22 February 1988.

1. Linthurst, R. A. *et al.* *Wat. Air Soil Pollut.* **31**, 577–591 (1986).
2. Jeffries, D. S. *et al.* *Wat. Air Soil Pollut.* **31**, 551–567 (1986).
3. Wright, R. F. & Henriksen, A. *Limnol. Oceanogr.* **23**, 487–498 (1978).
4. Wright, R. F. & Snekvik, E. *Verh. int. Verein. theor. angew. Limnol.* **20**, 765–775 (1978).
5. Norwegian State Pollution Control Authority Rep. 283/87 (1987).
6. Tang, A., Chan, W. H., Chung, D. H. S. & Lusi, M. A. Ontario Ministry of Environment Rep. ARB-008-86-AQM (1986).
7. Jeffries, D. S. in *Environmental Impacts of Smelters* (ed. Nriagu, J.), 117–154 (Wiley, New York, 1984).
8. Dillon, P. J. in *Environmental Impacts of Smelters* (ed. Nriagu, J.), 283–347 (Wiley, New York, 1984).
9. Keller, W. & Pitblado, J. R. *Wat. Air Soil Pollut.* **29**, 285–296 (1986).
10. Conroy, N. *et al.* in *Proc. 1st Speciality Symp. Atmos. Contribution Chem. Lake Wat. Internat. Assoc. Great Lakes Res.*, 28 Sept.–1 Oct., 1975.
11. Dillon, P. J., Jeffries, D. S. & Scheider, W. A. *Wat. Air Soil Pollut.* **18**, 241–258 (1982).
12. Dillon, P. J., Reid, R. A. & Girard, R. *Wat. Air Soil Pollut.* **31**, 59–65 (1986).
13. Kurtz, J. & Scheider, W. A. *Atmos. Environ.* **15**, 1111–1116 (1981).
14. Scheider, W. A., Jeffries, D. S. & Dillon, P. J. *Atmos. Environ.* **15**, 945–956 (1981).
15. Oliver, B. G., Thurman, E. M. & Malcolm, R. L. *Geochim. cosmochim. Acta* **47**, 2031–2035 (1983).
16. Cook, R. B. & Schindler, D. W. *Envir. Biogeochem. Ecol. Bull.*, Stockholm **35**, 115–127 (1983).
17. Kelly, C. A. *et al.* *Biogeochemistry* **3**, 129–140 (1987).
18. Harvey, H. H. *et al.* *Natn. Res. Council Can. Publ. No.* 18475 (1981).
19. Schindler, D. W. *et al.* *Science* **228**, 1395–1401 (1985).
20. Mierle, G., Clark, K. & France, R. *Wat. Air Soil Pollut.* **31**, 593–604 (1986).
21. Federal/Provincial Research and Monitoring Co-ordinating Committee, Canada. *Assessment of the State of Knowledge on the Long-range Transport of Air Pollutants and Acid Deposition*, part 3 (1986).

Detection of phosphine: new aspects of the phosphorus cycle in the hydrosphere

István Dévai*, Lajos Felföldy†, Ilona Wittner* & Sándor Plósz‡

* Hajdu-Bihar County Water and Canalization Works, PO Box 10, H-4001 Debrecen, Hungary

† Department of Ecology, L. Kossuth University, Egyetem tér 1, H-4010 Debrecen, Hungary

‡ Hungarian Institute for Town and Regional Planning, PO Box 20, H-1253 Budapest, Hungary

The role of phosphorus in limiting organic matter production and in causing eutrophication has been in the forefront of hydrobiological research during the past 30–50 yr^{1–7}. Comparing the cycles of major biogenic elements, it is evident that, with the exception of phosphorus, each of them contains gaseous substances (for example, CO₂, CH₄, O₂, N₂, NH₃, H₂S and volatile organic sulphur compounds) which, because of their gaseous state, can leave aquatic systems^{8–16}. We examined the phosphorus cycle of open-air sewage treatment plants and a deficit (~30–45%) in the phosphorus mass balance was found which cannot be explained by knowledge based on earlier research on this cycle of the hydrosphere. By developing special sampling and analytical methods, we have shown that gases released from the sewage treatment plants and from the sediments of shallow (1–2 m deep) waters contain a reduced, gaseous phosphorus compound: phosphine. According to our measurements and calculations, about 5 g of phosphorus per day was released as phosphine from an Imhoff tank settling 2,000 m³ per day of raw sewage. Under laboratory conditions, it was also demonstrated that phosphine is released by bacterial reduction from a medium containing inorganic phosphorus. The phosphorus content of the medium decreases by nearly one half. Our results on the metabolic importance of phosphine formation and release suggest that ~25–50% of the phosphorus deficit in open-air sewage treatment plants can be explained by the release of phosphine into the atmosphere. These results change our understanding of the aquatic phosphorus cycle; former ideas about the phosphorus budget should be revised.

Table 1 Water and phosphorus budget in the two systems examined

	BMCO		EUTROSTOP	
	input	output	input	output
Water (10 ³ m ³ yr ⁻¹)	248*	248†	190‡	70‡
Total P (mg l ⁻¹)				
min.	6.08	2.78	0.07	0.01
max.	27.20	13.04	1.27	0.46
average	10.17	6.89	0.45	0.097
Standard deviation	±3.11	±1.95	±0.31	±0.103
n	62	58	168	168

* Average of 11.5 yr.

† Because the BMCO system at Püspökladány was constructed on coherent clay soil, there was practically no seepage (measured permeability coefficient, $k = 1.5 \times 10^{-8}$ cm s⁻¹).

‡ Measured values during 1984; the difference between water input and output, with the exception of the evaporation loss of 6×10^3 m³ yr⁻¹, is due to seepage ($190 \times 10^3 - 70 \times 10^3 - 6 \times 10^3$) 114×10^3 m³. This large seepage rate might be because the water surface of the lake is higher by 1.2 m than the surrounding land and the subsoil is not impermeable clay, but peaty meadow soil.

Investigation of the P-budget in various sewage treatment systems showed that the total phosphorus (P_t) content of effluent is always lower than that of raw sewage. For two systems we had such data at our disposal, which made accurate P-budget estimations possible. One of these was a biomechanical combined oxidation (BMCO) stabilization lake system^{17–19}. A BMCO with a nominal capacity of 1,000 m³ per day consists of nine basins: a central, circular basin of 4,000 m³ capacity, lined with concrete slabs and aerated artificially with rotors, and eight earth basins (two of 8,000 m³ and six of 3,000 m³ capacity), built in double concentric rows around the central basin. The other system was a lake (EUTROSTOP system, no. 174,397, Hungarian patent) constructed to remove plant nutrients (phosphorus and nitrogen compounds) from polluted streams by means of aquatic vegetation. The experimental lake first constructed for this purpose was a rectangular earth basin 143 m long and 70 m wide, with an average water depth of 1.04 m. The lake abounded in macrophytes, and ~20% of its surface was densely covered with reed and bulrush stands. The water of the stream fed the basin at 10 points along its narrower side, and was drained at 10 surface points on the basin's opposite side.

The P-budget for the two systems is summarized in Tables 1 and 2. These data clearly point to an essential difference between 'income' (P_i in the influent) and 'expenditure' (P_e leaving the system through actual and potential sinks) in the P-budgets of the two systems (Table 2). The unaccounted for P for BMCO was 678 kg yr⁻¹, i.e. nearly 30% (26.9%) of the P-load. P-loss for EUTROSTOP (37.9 kg yr⁻¹) represents a higher value than that of the P-load of the system (44.3%).

Previous investigations on the sulphur cycle in shallow waters contributed significantly to our understanding of this anomaly in the P-balance^{20–22}. By analogy with the biochemical reduction of sulphate to hydrogen sulphide and nitrate to ammonia, reduction of phosphate to phosphine was considered feasible. Phosphine is a malodorous, colourless, poisonous gas. Its boiling temperature is -87 °C. At 17 °C, 26 cm³ phosphine dissolves in 100 cm³ water, but under natural conditions this value depends on partial pressure. Pure phosphine only ignites when heated to >100 °C. The simplest method of phosphine formation is the decomposition of alkali-metal or alkali-earth-metal phosphides with water²³. Phosphine, like hydrogen sulphide, is likely to be the final result of a multistage biochemical reaction. Its formation, in our opinion, cannot be either explained or ruled out by the properties of inorganic redox systems. The redox potential measurable in waters and sediments depends only on the actual dominant process determining electrode potential (with the platinum electrode only the potential of the redox system that

has the highest electron exchange current is measured^{24,25}). Further, redox potential values in the environment should not be compared with those of enzymatically catalysed pathways. Mainly on the basis of experience with sulphur compounds, it is thought most probable that other gaseous, reduced P-compounds (for example, diphosphine, volatile organic P-compounds) can also form, but there is no verification of this assumption. It has been stated (for example, by Iverson *et al.*²⁶) that, during the anaerobic corrosion of iron, sulphate-reducing bacteria produce volatile P-compounds by the reduction of inorganic P. To test the hypothesis that gases released from sediments of sewage plants and shallow lakes contain phosphine, conventional analytical, gas chromatographic and mass spectrographic methods were used.

Conventional analysis is based on the principle that phosphine (and other gaseous reduced P-compounds) can be transformed into orthophosphate and subsequently determined by photometry. The main limitation of this method is that results are not compound-specific. Such analysis was performed at the sewage plant of Hajdúböszörmény. The gas mixture (containing 65–70% methane) released from the Imhoff tank was transferred through a rotameter into a gas blowing burner with simultaneous introduction of oxygen and ignited. Gaseous, reduced P-compounds were oxidized, and flue gases were quantitatively absorbed in reagent grade sodium hydroxide²⁷. The amount of phosphate determined following its dissolution in distilled water by the usual colorimetric method^{28,29} was equivalent to 1.213 mg phosphorus. The quantity of burned gas, determined from the time of burning (50 min) and the value adjusted on the rotameter ($770 \text{ cm}^3 \text{ min}^{-1}$), was $50 \times 0.77 = 38.5 \text{ dm}^3$. With these values, computed phosphine content of gases released in the Imhoff tank was $1.213 \text{ mg} / 38.5 \times 10^{-3} \text{ m}^3 = 31.5 \text{ mg P m}^{-3}$.

Technical experiences in Hungary have shown that, for average town sewage treatment plants, the putrefaction of primary sludge in the Imhoff tank produces $50\text{--}100 \text{ l gas m}^{-3} \text{ d}^{-1}$ (ref. 30). Considering an average value ($75 \text{ l m}^{-3} \text{ d}^{-1}$), 4.725 g P ($0.075 \times 2,000 \times 0.0315 = 4.725$) per day leaves the putrefactive space of the Imhoff tank of $2,000 \text{ m}^3$ per day capacity at Hajdúböszörmény in gaseous form.

Quantitative determinations of phosphine, particularly its qualitative identification, can be performed with much greater accuracy by gas chromatography. Because in gas chromatography, it is theoretically possible for the various gaseous compounds to appear with identical or relatively identical retention times³¹, we used five types of columns and two detectors. Both detectors (flame photometric detector, FPD, and nitrogen-phosphorus flame ionization detector, N-P FID) were specifically sensitive to P-compounds. The absolute retention times of phosphine are summarized in Table 3.

Mass spectrometry is one of the most reliable methods of phosphine identification, and was used in the study of the BMCO sewage plant at Hajdúböszörmény. Measurements were made with a VG-7035 instrument, with 'VG High Resolution Software'. Samples were injected into the instrument as toluene solution treated with potassium hydroxide. We verified by high-resolution measurements, that the exact mass value of the peak of 34 in the peak appearing with a retention time of 1.15 min was 33.9969, which can be identified as phosphine.

These results verified that the gases released from the sediments of sewage plants and shallow lakes contain phosphine. The importance of phosphine formation and release can be judged only if phosphine and then the P-budget of the systems can be quantified. Using the gas chromatographic method (no. 2 in Table 3), 112 mg m^{-3} phosphine was measured on average (min. 11.6; max. 382; s.d. ± 103.64 ; $n = 11$). For quantitative estimation we must also know the rate of gas release. Measurements in the most polluted earth basin (B basin) of the BMCO showed that rate of gas release ranged from 8 to $241 \text{ l h}^{-1} \text{ m}^{-2}$. When pressing the sediment several times with an oar (to assess a stronger gas release induced by a sudden air depression), this

Table 2 Phosphorus budget in BMCO and EUTROSTOP

	Total phosphorus content (kg yr^{-1})	
	BMCO	EUTROSTOP
A Total P in influent	2,522	85.5
B Loss through seepage	1	18.2*
C In submerged macrophytes (October)	0	5.9
D In the plants of bulrush stands (October)	20	16.4
E Loss by chironomid emergence	5	0.3
F Accumulated in the sediment	113†	(−36.7)‡
G Loss through effluent	1,705	6.8
H Phosphorus deficit, $H = A - (B + C + D + E + F + G)$	678	37.9

A and G: calculated from data in Table 1 (cubic metres of water $\times$ average P_i kg m^{-3}).

C and D: total P-content in plants was estimated as follows. The surface area of the basin was divided into six areas with homogeneous vegetation. From every area, plants present in five water columns of 1 m^2 ground surface each were collected, and the fresh phytomass weighed, supposing 10% dry matter and 0.3% P in submerged aquatics, and 60% dry matter and 0.2% P in the bulrush^{36,37}.

E: P removed by chironomids was calculated on the basis of Lake Balaton data: 700 individuals $\text{m}^{-2} \text{ yr}^{-1}$ emergence corresponds to $0.03 \text{ g P m}^{-2} \text{ yr}^{-1}$ (ref. 38). In BMCO higher densities were considered (five generations, with density of 1,200 specimens per square metre).

* Average P-content of percolating water was estimated monthly on the basis of the average (0.16 mg l^{-1}) of all P-measurements made in lake water at 10 m distances along the longitudinal axis ($114 \times 10^3 \text{ m}^3 \text{ yr}^{-1} \times 0.16 \times 10^{-3} \text{ kg m}^{-3} = 18.2 \text{ kg P yr}^{-1}$).

† Five sediment samples were collected in 15 cm plastic tubes from each earth basin in the BMCO. Following freezing, the sediment samples were divided into three parts of 5 cm length. Each part was homogenized and analysed for its P-content^{28,29}. Thus on the basis of the analysis of ($8 \times 5 \times 3$) 120 samples, P-accumulation was calculated for each basin, and the sum of accumulations (1,300 kg) was divided by the number of years ($1,300 \text{ kg} / 11.5 \text{ yr} = 113 \text{ kg yr}^{-1}$).

‡ Analysis of 12 sediment samples showed P-content in the upper 5 cm sediment layer to be 102 kg ha^{-1} . Analysis of soil samples from the area surrounding the lake before its construction yielded a P-content of $138.7 \text{ kg P ha}^{-1}$, which is more suggestive of diminution than accumulation ($102 - 138.7 = -36.7 \text{ kg P ha}^{-1}$). This diminution might be explained by the leaching effect of percolating water. If this was not so, the P-loss for EUTROSTOP in this table should be increased by 36.7 kg. Because of this uncertainty, no allowance was made for this value in the estimation of the phosphorous deficit.

value increased to $84 \text{ l h}^{-1} \text{ m}^{-2}$. On the basis of measured values, the rate of average yearly gas release proved to be $213 \text{ m}^3 \text{ yr}^{-1} \text{ m}^{-2}$ (min, 70; max, 736; s.d. ± 220 ; $n = 8$). Because in the other basins of the BMCO system, gas release was less intensive, only one third ($8,000 \text{ m}^2$) of the total surface ($22,500 \text{ m}^2$) was considered in the calculations. Estimated annual gas release from the BMCO was ($213 \times 8,000$) $1.7 \times 10^6 \text{ m}^3 \text{ yr}^{-1}$ on average (min, 5.6×10^5 ; max, 5.9×10^5). If the phosphine concentration and the rate of gas release are known, loss into the atmosphere can be calculated. In the BMCO system, the loss was estimated to be $1.12 \times 10^{-4} \text{ kg m}^{-3} \times 1.7 \times 10^6 \text{ m}^3 \text{ yr}^{-1} = 190 \text{ kg yr}^{-1}$ phosphine on average, and the same computed for $P \sim 175 \text{ kg yr}^{-1}$ (min. $\times$ min. = $6.5 \text{ PH}_3 \text{ kg yr}^{-1}$; max. $\times$ max. = $2,254 \text{ PH}_3 \text{ kg yr}^{-1}$).

Gas release in EUTROSTOP at Badacsony-tomaj was measured at four points on the basin under normal circumstances (average, $0.47 \text{ l h}^{-1} \text{ m}^{-2}$) and after pressing the sediment several times with an oar (average, $515 \text{ l h}^{-1} \text{ m}^{-2}$). The large difference between the two averages suggest that here, unlike for BMCO, gas release is not continuous but intermittent and follows, for example, a sudden air-pressure drop. For the estimation of yearly gas release, the amount of sudden air-pressure drop was taken as 24 h (i.e. 1 day out of 365 days). Thus, on

Table 3 Main gas chromatographic parameters of phosphine determination

Number	Detector	Column	Temperature (°C)			Retention time of phosphine (min)
			Detector	Injector	Oven (isotherm)	
1	N-P FID	1.8 m SS, 1.5% OV-17/1.95% OV-210 on 100/120 Supelcoport	300	150	50	0.45
2	N-P FID	1.8 m SS, Chromosorb 103 (ref. 39)	300	100	75	1.43
3	FPD (ref. 40)	1.8 m SS, Chromosorb 103 (ref. 39)	200	100	75	1.43
4	N-P FID	25 m × 0.32 mm × 0.17 µm HP-Ultra 1*	300	120	50	1.47
5	N-P FID	1.8 m SS, Porapak Q	300	150	75	2.03
6	FPD (ref. 40)	1.2 m Teflon, Chromosil 330	100	80	0	0.90

The gas mixture from sediments was transported into the laboratory in septate glass bottles or in toluene solution. Aliquots were directly injected into the gas chromatograph. These investigations started in 1984 were performed with Hewlett-Packard 5840 Å gas chromatograph-9825 T calculator system, applying nitrogen of high purity as the carrier gas.

* In GC/MS analysis the same column was used (helium as carrier gas, EI operation mode, 70 eV ionization energy).

the basis of measured values, estimated gas release from EUTROSTOP of 1 ha surface area is $(0.47 \times 10^{-3} \times 24 \times 364 \times 10,000 + 0.515 \times 24 \times 10,000) 1.65 \times 10^5 \text{ m}^3 \text{ yr}^{-1}$. Consequently P-loss into the atmosphere can be estimated as $1.12 \times 10^{-4} \text{ kg m}^{-3} \times 1.65 \times 10^5 \text{ m}^3 \text{ yr}^{-1} = 18.48 \text{ kg yr}^{-1}$ phosphine, which is equivalent to $16.85 \text{ kg P yr}^{-1}$. This value, as expected, is lower by one order than the average for BMCO. Processes of material turnover are more intensive in the sewage plant.

In Table 2, about 25% (175 kg yr^{-1}) of the unaccounted for P-deficit for BMCO (678 kg P yr^{-1}) and about 50% (16.85 kg yr^{-1}) of the same for EUTROSTOP ($37.9 \text{ kg P yr}^{-1}$) can be explained by the release of phosphine.

We attempted to verify phosphine formation *in vitro*, despite the fact that there were no references to this either in hydrobacteriological or other literature³²⁻³⁵. We detected phosphine by gas chromatography from the air space of an anaerobic bacterium culture grown on a medium containing organic substances (Lab-Lemco, Peptone), inorganic P-compounds ($\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, K_2HPO_4) and other substances (NH_4Cl , MgSO_4 , CaCO_3 , filter paper, tap water). The P-concentration of a continuous bacterium culture containing 340 mg P l^{-1} medium decreased to 198 mg P l^{-1} during 56 days, showing that microorganisms can generate large amounts of phosphine from phosphate. These results are important because they support the foregoing arguments on phosphine formation and release in sewage treatment plants under aquatic vegetation. They also point to the prospects of investigating *in vitro* phosphine production in future studies.

We thank Dr Gy. Dévai for his continuous help and for his Chironomidae studies, Dr Z. Dinya for the mass spectrometric measurements, and Dr B. Harman for help in the gas chromatographic investigations. We acknowledge the referees for useful comments.

19. Dévai, I. & Woynarovich, E. in *Some Approaches to Saprobial Problems* (ed. Sudzuki, M.) 37-47 (Sanseido, Tokyo, 1981).
20. Dévai, I. *et al. Envir. Pollut.* **B8**, 155-160 (1984).
21. Dévai, I., Dévai, Gy. & Wittner, I. *Arch. Hydrobiol. Monogr. Beiträge* **4**, 534-579 (1985).
22. Heim, Cs., Dévai, I. & Harangi, J. *J. Chromatogr.* **295**, 259-263 (1984).
23. *Gmelins Handbuch der anorganischen Chemie. Phosphor*. Part C (Verlag Chemie, Weinheim, 1965).
24. Stumm, W. & Morgan, J. J. *Aquatic Chemistry. An Introduction Emphasizing Chemical Equilibria in Natural Waters* (Wiley-Interscience, New York, 1970).
25. David, R. *Radiat. Environ. Biophys.* **25**, 219-229 (1986).
26. Iverson, W. P., Olson, G. J. & Heverly, L. F. in *Biologically Induced Corrosion* (ed. Dexter, S. C.) 154-161 (Natl. Assoc. of Corrosion Engineers, Houston, Texas, 1985).
27. Dévai, I. *Acta Biol. Debrecina* (in the press).
28. Murphy, J. & Riley, J. P. *Analytica Chim. Acta* **27**, 31-36 (1962).
29. Gales, M. E., Julian, E. C. & Kroner, R. C. *J. Am. Wat. Wks. Ass.* **58**, 1363-1368 (1966).
30. Horváth, I. *Mechanical Waste Treatment VIZDOK* (Budapest, 1973).
31. McNair, H. M. & Bonelli, E. J. *Basic Gas Chromatography* (Consolidated Printers, Berkeley, California, 1969).
32. Rheinheimer, G. *Mikrobiologie der Gewässer* (Gustav Fischer, Jena, 1971).
33. Daubner, I. *Mikrobiologie des Wassers* (Akademie, Berlin, 1972).
34. Rodina, A. G. *Methods in Aquatic Microbiology* (University Park Press, Baltimore, 1972).
35. Gorlenko, V. M., Dubinina, G. A. & Kuznetsov, S. I. *The Ecology of Aquatic Micro-organisms* (E. Schweizerbart, Stuttgart, 1983).
36. Tóth, L. *Res. Water Quality Technol., Budapest* **2**, 48-55 (1972).
37. Kovács, M. & Tóth, L. *Proc. VITUKI, Budapest* **14**, 49-74 (1979).
38. Dévai, Gy. in *Chironomidae* (ed. Murray, D. A.) 269-273 (Pergamon, Oxford, 1980).
39. Bossard, A., Kamga, R. & Raulin, F. *J. Chromatogr.* **330**, 400-402 (1985).
40. Hashimoto, S., Fujiwara, K. & Fuwa, K. *Analyt. Chem.* **57**, 1305-1309 (1985).

Attenuating organic contaminant mobility by soil modification

Stephen A. Boyd, Jiunn-Fwu Lee & Max M. Mortland

Department of Crop and Soil Sciences, Michigan State University, East Lansing, Michigan 48824, USA

Low organic matter clays, soils and aquifer materials have very little sorptive capability for common groundwater contaminants. Here we use organic cations of the form $[(\text{CH}_3)_3\text{NR}]^+$ to displace naturally occurring exchange ions resulting in significantly higher organic matter contents, and greatly enhanced sorptive properties for nonionic organic solutes. The organic phase derived from exchanged hexadecyltrimethylammonium, where R is a C-16 hydrocarbon, behaved as a powerful partitioning medium that was 10 to 30 times more effective on a unit weight basis than natural soil organic matter for removing benzene, dichlorobenzene and perchloroethene from water. This simple soil modification might be used to improve the retardation capabilities of low organic matter soils and aquifer materials, and to enhance the containment capabilities of clay landfill liners and bentonite slurry walls.

The sorption of non-ionic organic contaminants (NOCs) from water by soil is controlled by the soil organic matter content^{1,2}. In contrast, the soil mineral fraction is generally not important in the uptake of NOCs in soil-water systems. Dipole interactions between the mineral surfaces and water prohibit NOCs from interacting with this fraction of soil³. One practical result of the

Received 22 June 1987; accepted 7 March 1988.

1. Sawyer, C. N. *Sewage ind. Wastes*, **24**, 768-776 (1952).
2. Ohle, W. *Vom Wasser* **20**, 11-23 (1953).
3. Lund, J. W. *Nature* **214**, 557-558 (1967).
4. Thomas, E. A. *Mitt. int. Verein. theor. angew. Limnol.* **14**, 231-242 (1968).
5. Thomas, E. A. *Wass. u. Energ. Wirt.* **60**, 115-125 (1968).
6. Edmondson, W. T. *Science*, **169**, 690-691 (1970).
7. Vollenweider, R. A. & Dillon, P. J. *The Application of the Phosphorus Loading Concept to Eutrophication Research* (Nat. Res. Council of Canada, Ottawa, 1974).
8. Hutchinson, G. E. *A Treatise on Limnology I* (Wiley, New York, 1957).
9. Reid, G. K. *Ecology of Inland Waters and Estuaries* (Van Nostrand, New York, 1961).
10. Dussart, B. *Limnologie* (Gauthier-Villars, Paris, 1966).
11. Wetzel, R. G. *Limnology* (Saunders, Philadelphia, 1975).
12. Svensson, B. H. & Söderlund, R. (eds) *Nitrogen, Phosphorus and Sulphur - Global Cycles. SCOPE Report 7. Ecol. Bull., Stockholm*, **22**, 1-192 (1976).
13. Emsley, J. in *The Handbook of Environmental Chemistry Vol. 1, Part A* (ed. Hutzinger, O.) 147-167 (Springer, Berlin, 1980).
14. Goltzman, H. L. & Kouwe, F. A. in *The Functioning of Freshwater Ecosystems IBP 22* (eds Le Cren, E. D. & Lowe-McConnell, R. H.) 85-140 (Cambridge University Press, 1980).
15. Moss, B. *Ecology of Fresh Waters* (Blackwell Scientific, Oxford, 1980).
16. Uhlmann, D. *Hydrobiologie* (Gustav Fischer, Jena, 1982).
17. Dévai, I. *Acta Biol. Debrecina* **14**, 51-65 (1977).
18. Dévai, I. *Acta Biol. Debrecina* **14**, 67-78 (1977).

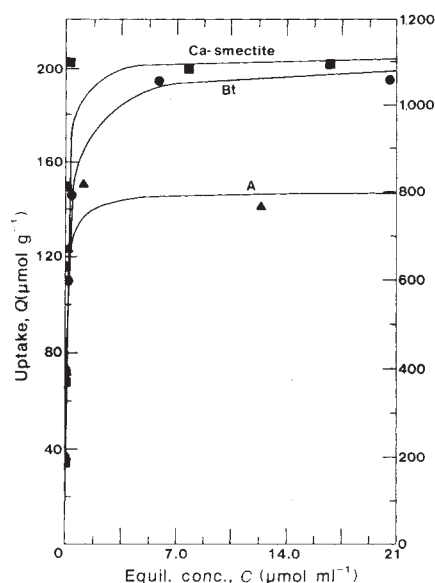


Fig. 1 Adsorption of hexadecyltrimethylammonium (HDTMA) iodide (methyl- ^{14}C , American Radiolabeled Chemicals, Inc., specific activity 52 mCi mmol $^{-1}$) by Ca-smectite (right scale) and Marlette soil A and Bt horizons (left scale).

Methods. Either 0.2 g Ca-smectite or 1 g soil and 10 ml water were placed in glass centrifuge tubes. Portions (50–400 μl) of an ethanol solution containing 1:1000 ^{14}C -HDTMA:HDTMA were added in an amount equal to 0.25, 0.5, 0.75, 1, 1.5 and 2 times the cation exchange capacity (CEC) of the clay or soil. The samples were shaken 24 h at 25 $^{\circ}\text{C}$. The aqueous phase was separated by centrifugation and 1 ml assayed for ^{14}C by liquid scintillation counting.

strong interaction of water with clays is the lack of sorptive capability of clay containment barriers. Similarly, low organic matter surface soils and aquifer materials have very little retardation capability for NOCs. As a result, many organic chemicals are found as groundwater contaminants and represent a serious threat to human health.

Soil materials contain significant cation-exchange capacity (CEC) derived primarily from clay minerals. Clays possess a net negative electrical charge which is compensated for by exchange cations (such as Ca^{2+}) on their surfaces. Hydration of these metal cations imparts a hydrophilic nature to the mineral surfaces. Organic cations may enter into ion exchange with metal cations on the exchange sites of clays. With organic cations that contain sizable organic moieties, the surface of the clay may be greatly modified to become strongly organophilic.

We have previously described the sorptive characteristics of organo-clay complexes formed with purified smectite clays^{4,5}

and organic cations of the form $[(\text{CH}_3)_3\text{NR}]^+$. Here we present data showing the enhanced uptake of several mobile organic contaminants from water by actual soil materials exchanged with hexadecyltrimethylammonium $[(\text{CH}_3)_3\text{N}(\text{CH}_2)_{15}\text{CH}_3]$ (HDTMA) cations.

Adsorption isotherms of ^{14}C -HDTMA on Ca^{2+} -smectite and the Marlette soil are shown in Fig. 1. The adsorption of HDTMA is very strong as expected. These are type-I isotherms characteristic of ion exchange reactions in which the added ion is strongly preferred. Figure 1 shows that HDTMA is very strongly sorbed from solution until the CEC of the clay or soil material (Table 1) is exceeded. The large alkyl ammonium ions once bound are very difficult for smaller inorganic ions to displace due to their high mass and because of van der Waals' interactions between the alkyl chains themselves⁶. As a result, large organic cations like HDTMA bind almost irreversibly to the clay surface.

Table 1 shows properties of the untreated and HDTMA-treated soil materials. The treated materials were prepared by adding HDTMA in an amount equivalent to the CEC of the clay or soil. It is obvious from Table 1 that the organic matter contents have increased significantly. The measured organic-matter contents agree well with the expected amount, again showing the stoichiometric displacement of inorganic ions by HDTMA.

Typical soil sorption behaviour was observed for the untreated soil materials. For benzene and perchloroethene, uptake from water was very low by the untreated Bt horizon, due to its low organic matter content (0.6%), and could not be distinguished from the blank (no soil) samples (Fig. 2). Uptake by the untreated A horizon was higher because of the higher organic matter content (5.2%), and the sorption isotherms were highly linear. Uptake of NOCs from water by soil is commonly expressed as $Q = KC$. The sorption coefficients K , normalized for the soil organic matter (om) content ($K_{\text{om}} = 100K/\% \text{om}$), tend to converge on a single value for each solute^{1,2}. The log K_{om} values obtained here using the A-horizon soil are very close to those reported for other surface soils².

After treating the same soil materials with HDTMA in an amount equivalent to the CEC, the sorption of benzene, dichlorobenzene and perchloroethene increased dramatically. This effect is seen in Fig. 2, which shows the sorptive capability of the untreated and HDTMA-treated Bt horizon soil. For the A horizon the K values increased by a factor of 12–16 in the HDTMA-treated material (Table 2). The increase in K was even greater for the Bt horizon which, in the untreated form, contains very little natural organic matter. For dichlorobenzene, K was about 200 times greater in the modified Bt soil.

An estimate of the relative effectiveness of organic matter derived from HDTMA versus natural organic matter can be made by comparing the log K_{om} values from the untreated A horizon and from the HDTMA-treated Bt horizon. In the former

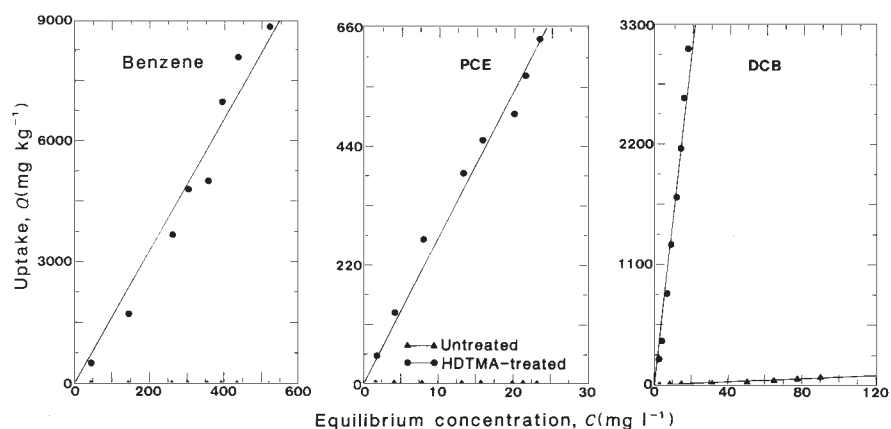
Table 1 Properties of the untreated and HDTMA-treated smectite clay and Marlette soil

Property	Ca-smectite		Bt horizon		A horizon	
	Untreated	HDTMA-treated	Untreated	HDTMA-treated	Untreated	HDTMA-treated
CEC (meq/100 g)	91		14.6		16.4	
Particle size (%)						
Sand	0		42.6		56.6	
Silt	0		16.7		22.0	
Clay	100		40.7		21.4	
Organic carbon (%)	0.91	17.3	0.3	3.7	2.6	6.5
Organic matter (%)						
Soil organic matter	1.8	1.8	0.6	0.6	5.2	5.2
HDTMA organic matter	0	20.5	0	4.25	0	4.9
Total organic matter	1.8	22.3	0.6	4.85	5.2	10.1

Cation exchange capacity (CEC) is sum of extractable bases: $\text{Al} + \text{Ca} + \text{Mg} + \text{K} + \text{Na}$ (ref. 7). Particle size analysis was performed by the Michigan State University Soil Testing Laboratory. Organic carbon analysis was performed by Huffman Laboratories, Inc., Golden Colorado. Factors used to convert soil and HDTMA organic carbon to organic matter were 2 and 1.2, respectively. The Marlette soil is classified as a fine-loamy, mixed Glossoboric Hapludalph. The clays in this soil are mainly illitic with minor components of chlorite and kaolinite.

Fig. 2 Sorption of benzene, perchloroethene (PCE) and 1,2-dichlorobenzene (DCB) by untreated and HDTMA-treated Marlette soil Bt horizon.

Methods. Treated soil was prepared by mixing 200 g soil and 1800 ml water, and then adding 50 ml H₂O containing HDTMA in an amount equal to the CEC of the soil. After thorough mixing, the mixture was filtered, washed with water and freeze dried. For the isotherms, 3 g treated soil or 15 g untreated soil, and 25 ml water were placed in glass centrifuge tubes. Three to 45 μ l of benzene (neat), or of 0.1 g ml⁻¹ methanol solutions containing PCE or DCB, were added. The tubes were closed immediately and shaken 24 h at 25 °C. The aqueous phase was separated by centrifugation and either analysed directly for benzene by HPLC with UV-detection or extracted with hexane and this extract analysed for PCE or DCB by GC with electron-capture detection.



case, organic matter is entirely natural, and in the latter organic matter is derived primarily from HDTMA (Table 1). This comparison reveals that the organic matter composed of HDTMA is about 10–30 times more effective than natural soil organic matter for removing the organic solutes from water (Table 2). Thus, not only can the organic matter content of soil be increased through modification with HDTMA, but the added organic matter is highly effective for immobilizing common mobile pollutants like benzene.

The organic phase formed by HDTMA acts as a solubilizing (partitioning) medium for removing NOCs from water⁴. This would be functionally and conceptually similar to a bulk organic solvent like hexane or octanol. This concept is supported by the linearity of the isotherms^{1,2} and by the similarity of the log K_{om} values (from the HDTMA-treated Bt) to the corresponding octanol–water or hexane–water partition coefficients (Table 2). The organic phase derived from exchanged HDTMA is more effective than soil organic matter for removing NOCs from water because it is derived entirely from the low-polarity C-16 hydrocarbon tails whereas soil organic matter contains a high polar group content.

Our results demonstrate that HDTMA-exchanged soil materials and clays are effective sorbents for removing common organic groundwater pollutants like benzene and perchloroethene from water. The soil modification is accomplished by simple ion exchange of highly preferred large organic cations like HDTMA for small inorganic cations. Quaternary organic cations like HDTMA are common cationic surfactants and available commercially. Such modified soils and clays may be used: (1) to increase the sorptive capacity of low organic matter soils; (2) to increase the sorptive properties of subsurface materials under existing waste disposal sites via underground injection of the organic cation; and (3) as components of ben-

tonite slurry walls and clay landfill liners to enhance the containment capabilities of such waste disposal reservoirs. The physical mixing of or layering of HDTMA–bentonite and dispersed Na-bentonite would result in a composite having the two important characteristics of impermeability to water, and strong sorptive capabilities. Although many engineering aspects remain to be addressed in the actual field application of this technology, it is clear that the materials described here display strong sorptive capabilities and may therefore provide an effective retardant for the movement of organic contaminants.

We thank C. T. Chiou for consultations and J. Crum for help in obtaining soil samples. Partial support from US Environmental Protection Agency (Cincinnati) is gratefully acknowledged. Journal article no. 12593 of the Michigan Agricultural Experiment Station.

Received 30 November 1987; accepted 28 March 1988.

1. Chiou, C. T., Peters, L. J. & Freed, V. H. *Science* **206**, 831–832 (1979).
2. Chiou, C. T., Porter, P. E. & Schmedding, D. W. *Environ. Sci. Technol.* **17**, 227–231 (1983).
3. Chiou, C. T. & Shoup, T. D. *Environ. Sci. Technol.* **19**, 1196–1200 (1985).
4. Boyd, S. A., Mortland, M. M. & Chiou, C. T. *Soil Sci. Soc. Amer. J.* **52**, 652–657 (1988).
5. Boyd, S. A., Shaobai, S., Lee, J. F. & Mortland, M. M. *Clays Clay Min.* **36**, 125–130 (1988).
6. Theng, B. K. G. *The Chemistry of Clay–Organic Reactions*, 221–238 (Wiley, New York, 1974).
7. *Soil Survey Laboratory Methods and Procedures for Collecting Soil Samples* Report No. 1, 66 (US Dept. Agr. Soil Conservation Service, Washington, DC, 1972).
8. Hansch, C. & Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley, New York, 1979).

A numerical model of landform development by glacial erosion

Jonathan M. Harbor*, Bernard Hallet* & Charles F. Raymond†

* Department of Geological Sciences and Quaternary Research Center, University of Washington AJ-20, Seattle, Washington 98195, USA

† Geophysics Program and Quaternary Research Center, University of Washington AK-50, Seattle, Washington 98195, USA

Table 2 Sorption coefficients of benzene, perchloroethene (PCE) and 1,2-dichlorobenzene (DCB) on untreated and HDTMA-treated soil

Sorption coefficient	A horizon		Bt horizon		log K_{ow} *
	Untreated	HDTMA-treated	Untreated	HDTMA-treated	
K					
Benzene	0.57	7.8	0	16.5	
PCE	2.95	37.0	0	27.3	
DCB	7.25	120.0	0.80	162.6	
log K_{om}					
Benzene	1.04	1.89	—	2.53	2.3†
PCE	1.75	2.56	—	2.75	2.6
DCB	2.14	3.07	2.12	3.52	3.38

* Octanol–water partition coefficients².

† Hexane–water partition coefficient⁸.

The jagged peaks, knife-edged ridges, and steep-sided valleys of alpine landscapes are well known products of glaciation, yet little is known about how the dynamics of ice flow and glacial erosion give rise to such landforms. By linking a finite-element model for ice flow through a valley with a model for glacial erosion, we have simulated the gradual evolution of one of the most striking of glacial landforms, the U-shaped valley. Incorporating realistic glacier sliding laws and spatial variations in rock resistance, our simulations illustrate how ice flow, erosion patterns and topography interact in the long term evolution of glacial landforms. In addition, our results indicate for the first time that the formation of a steady-state U-shaped valley takes approximately 10⁵ years.

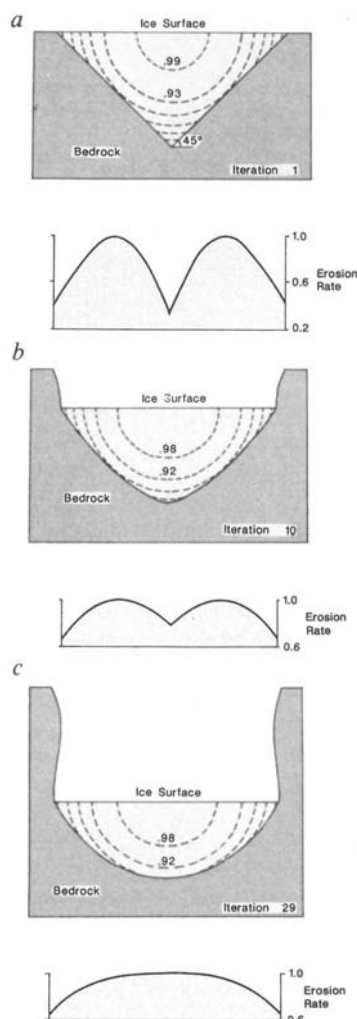


Fig. 1 Development of the cross-section of a glacial valley in homogeneous bedrock. Ice velocity and erosion rate values are normalized to cross-sectional maxima of 1.00, and the velocity contour interval is 0.06. *a*, Conditions associated with an initial V-shaped valley, *b*, an intermediate stage in form evolution, and *c*, the essentially steady-state form reached after lowering of the valley centreline by 90% of the initial valley depth. In subsequent iterations the glacial channel is simply translated downwards, leaving nearly vertical valley walls. Ice discharge and surface slope are constant over time, and ice flow is modelled as a standard power law material, with the strain rate proportional to the cube of the stress⁵. Glacier sliding is assumed to have the form described by Budd *et al.*⁶, and erosion normal to the bed at any point is taken to be proportional to the local sliding velocity squared^{2,7}.

In the past decade our ability to simulate the detailed characteristics of glacier and ice-sheet motion has been improved considerably by the development of accurate numerical ice-flow models that include realistic ice rheology and basal sliding (for example, ref. 1). At the same time, theoretical and empirical work in glacial geomorphology has provided new models of specific erosional processes^{2,3} that can now be used in the study of characteristic glacial landforms such as fjords, cirques and U-shaped valleys. Taking advantage of these recent advances in glaciology and glacial geomorphology, we have coupled a numerical ice-flow model with a physically based model of glacial erosion to study the development of typical alpine glacial-valley cross-sections.

As part of a pioneering study of landform evolution using ice-flow modelling, Oerlemans⁴ examined glacial-valley development under large ice-sheets. He showed that where much

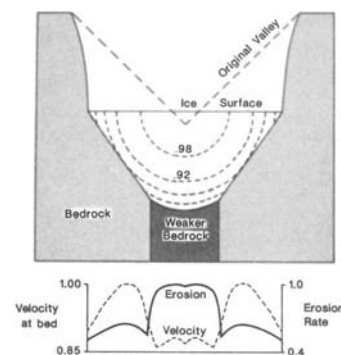


Fig. 2 Development of a glacial valley centred on relatively weak bedrock. Ice velocity, basal ice velocity, and erosion rate values are normalized to cross-sectional maxima of 1.00, and the velocity contour interval is 0.06. When the valley centreline has been lowered by 90% of the initial valley depth, a trough is conspicuous in the centre of the valley, where the bedrock is twice as erodible as that along the sides. However, the amount of overdeepening is limited because ice velocities, and thus erosion rates, are reduced as the flow is increasingly constrained in the developing trough.

of the base of an ice-sheet is frozen to its substrate, temperature patterns play a dominant role in valley-deepening as they dictate areas where basal sliding and thus glacial erosion occur. However, except in polar regions, these thermal influences are absent for typical valley glaciers, thus other glaciological processes must control cross-valley variations in erosion rate. Therefore any simulation of valley development for the case of alpine glaciers requires a detailed ice-dynamics model suitable for calculating cross-valley variations in ice flow for cases where basal sliding is widespread, coupled with a glacial erosion model.

In our simulations of valley development, we start by modelling glacier flow through an initial cross-section to calculate the glaciological parameters likely to control erosion: sliding velocity, basal stress, and rate of energy dissipation. We then numerically simulate erosion of the valley to produce a modified transverse profile, for which a new flow field and erosion pattern is computed. A number of iterations permit us to study the progressive transformation of the valley shape, which can be compared ultimately with published field data. In this approach, the iterative procedure is critical as it allows important interactions between topography and ice flow during landform evolution to be modelled accurately.

Our glacier dynamics model consists of a finite-element calculation of the pattern of flow for a power-law fluid in an inclined channel of arbitrary cross-section, assuming the flow is rectilinear. In this application we assume that the ice is at its melting point throughout, and that the strain rate is proportional to the cube of the stress, following Glen's flow law for ice⁵. The model can also accommodate a variety of boundary conditions, and in our initial simulations presented here we use a sliding law that relates sliding velocity (V) to basal shear stress (τ) and effective pressure (P_e), as $V \propto \tau^3 P_e^{-1}$ (ref. 6). The effective pressure is calculated assuming an effective water level at 80% of the ice thickness. This glaciological model is coupled with an erosion program that calculates local changes in bed elevation at regularly spaced intervals along the valley cross-section, on the basis of an erosion law that can use any of the glaciological parameters calculated by the finite-element program. In our initial simulations, erosion normal to the bed is taken to be proportional to the square of the local sliding velocity, in accord with Hallet's⁷ theoretical model for glacial abrasion and Shoemaker's⁷ arguments for glacial plucking.

In our first model run we start with a glacier occupying a typical V-shaped river valley (Fig. 1). The initial velocity solution for this case includes boundary maxima midway along the channel walls, which give corresponding peaks in the erosion

rate across the valley. This erosion pattern produces a curved channel cross-section beneath the ice, as anticipated in earlier work⁸, but as the channel shape changes, this alters both the velocity and erosion patterns. However, eventually the shape of the channel beneath the ice and the velocity field for the glacier remain essentially constant over time, although the glacier continues to incise vertically. This steady-state shape is roughly parabolic (U-shaped) and closely matches field measurements of the forms of glaciated valleys⁹⁻¹¹. Interestingly, reaching this steady-state required the valley to almost double its original depth, and this amount of incision would require $\sim 10^5$ years of continuous glacial erosion, given that glacial erosion rates are $\sim 10^{-3}$ m yr⁻¹ (ref. 12), and that glacial valleys are typically $\sim 10^2$ m deep.

A major asset of the numerical modelling approach is that idealized and observed distributions of rock resistance can be included to investigate the effects of geological inhomogeneities on the evolution of glacial landforms. In our second model run we consider glacial valley cross-profile development in a situation identical to the first, except that the central 30% of the valley consists of material twice as erodible as the surrounding rocks (Fig. 2). In this case, although the weaker material is preferentially eroded, the amount of local entrenchment is surprisingly limited. This occurs because the ice flow becomes constricted, and hence less erosive, in the trench that develops in the weaker rock. Clearly, the dominant role often ascribed to structural control deserves more detailed investigation.

Our coupling of a glaciological model with a simple erosion law allows simulations of topographic development that are consistent with observed glacial landforms. In addition, for the first time our simulations provide some means to estimate the time required for the development of specific glacial landforms, and allow an assessment of the importance of geological structures in controlling landform evolution. In future work, models of this type promise not only to improve our understanding of glacial landscapes, but may also provide a means to test alternative glacial erosion and sliding laws on the basis of their ability to simulate the development of known glacial landforms.

This research is being supported by the NSF and has benefited from discussions with Joseph Walder, Edwin Waddington, Robert Andersson and Neil Humphrey.

Received 3 December 1987; accepted 31 March 1988.

1. Hodge, S. M. *J. Glaciol.* **31**, 350-359 (1985).
2. Hallet, B. J. *Glaciol.* **17**, 209-221 (1979); *Ann. Glaciol.* **2**, 23-28 (1981).
3. Boulton, G. S. in *Glacial Geomorphology* (ed. Coates, D. R.) 41-87 (Binghamton, State University of New York, 1974).
4. Oerlemans, J. *Zeit. Gletscherk. Glazialgeol.* **20**, 107-126 (1984).
5. Glen, J. W. *Proc. R. Soc. A* **228**, 519-538 (1955).
6. Budd, W. F., Keage, P. L. & Blundy, N. A. *J. Glaciol.* **23**, 157-170 (1979).
7. Shoemaker, E. M. *J. Glaciol.* **32**, 65-71 (1986).
8. Johnson, A. *Physical Processes in Geology* (Freeman, San Francisco, 1970).
9. Svensson, H. *J. Glaciol.* **3**, 362-363 (1959).
10. Doornkamp, J. C. & King, C. A. *Numerical Analysis in Geomorphology: an Introduction* (St Martin's, New York, 1971).
11. Graf, W. L. *Arch. Alp. Res.* **2**, 303-312 (1970).
12. Drewry, D. *Glacial Geological Processes* (Arnold, London, 1986).

Tectonic and magmatic transitions in the Western Great Basin, USA

David S. Ormerod*, Christopher J. Hawkesworth*,
Nicholas W. Rogers*, William P. Leeman†
& Martin A. Menzies‡

* Department of Earth Sciences, Open University,
Milton Keynes MK7 6AA, UK

† Department of Geology, Rice University, PO Box 1892, Houston,
Texas 77001, USA

‡ Department of Geology, Royal Holloway and Bedford New
College, University of London, Egham TW20 0EX, UK

The isotope and trace-element geochemistry of late Cenozoic basalts from the Western Great Basin of California and Nevada demonstrates contributions from two distinct mantle source regions. High concentrations of large-ion lithophile relative to high-field strength elements ($50 < \text{Ba}/\text{Nb} < 250$) are often taken to reflect a contribution from contemporaneous subduction processes. But in this area they are associated with a marked change in $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ across a major lithospheric boundary, which is normal to the orientation of the palaeo-subduction zone and implies that these samples derived their isotope and trace element signatures from the mantle lithosphere. The second component is similar to that seen in ocean island basalts, and is attributed to asthenospheric source regions within the convecting upper mantle. Significantly, the onset of asthenosphere dominated magmatism migrated northwards during the Plio-Pleistocene about 2-3 Myr behind the trailing edge of the subducted slab. During this period it is proposed that asthenospheric diapirs rose from the level of the subducted slab into the zone of melt generation, implying ascent rates of between 5 and 8 cm per annum.

One of the major successes of plate tectonic theory is its ability to account for distinctive styles of magmatic activity in terms of lithospheric plate interactions. The intimate links between tectonic and magmatic processes are manifest both in the tectonic causes of melt generation and the consistent pattern of trace element ratios in basalts from different environments¹⁻⁵.

Not all the links are well understood, particularly the interplay between a tectonic process and the mantle source components which it renders accessible to magmagenesis. It is not yet clear, for example, to what extent the trace-element characteristics of lavas from convergent plate margins reflect a recycled sediment contribution via the subducted slab, or the physical process of melt generation in a hydrated mantle wedge⁶⁻⁸. A potentially fruitful approach is to constrain the nature of magmatic processes by determining the rate at which they respond to changes in tectonic regime. In continental areas, where this approach is facilitated by the greater range in isotope and trace-element ratios, such data would further constrain recent discussions of lithospheric versus asthenospheric contributions to continental magmatism. Here we report the results of an integrated isotope and trace-element study of late Cenozoic (<15 Myr) basaltic volcanism in the Western Great Basin (WGB) of California and Nevada in the United States (Fig. 1). An existing body of K-Ar age determinations (supplemented by this study) permits an examination of the temporal relationship between the magmatic and tectonic transitions that occurred during this period.

Late Cenozoic volcanism within the WGB includes alkali-olivine basalts (*ne-normative*), olivine basalts (*hy-normative*), basaltic andesites, and an occurrence of potassic lavas^{9,10} (representative analyses are presented in Table 1). This magmatic activity occurred within an extensional environment situated above a subduction zone, and spans a 10-Myr period during which slab subduction effectively ceased. The subduction of Pacific oceanic lithosphere ended diachronously beneath California and Nevada, beginning in the south at the end of the Oligocene and then progressing northwards during the Miocene as the San Andreas transform system developed¹¹. The transition from a supra-subduction zone to within-plate environment and the rate of magmatic response to removal of the subducted slab provides a direct constraint on models of intra-plate magmatism.

In keeping with a changing tectonic environment, the trace-element geochemistry of alkali basalts (all with MgO contents >4 weight per cent) from the WGB range between two distinct end-member types. At one extreme, samples show a smooth concave downwards trace-element pattern, with high concentrations of incompatible relative to more compatible elements, and

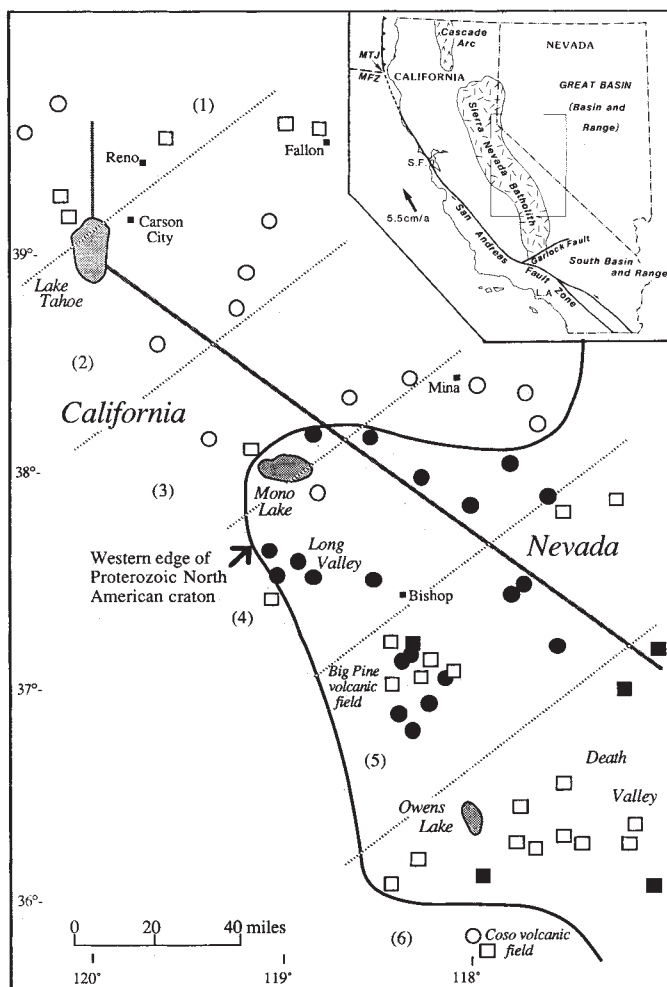


Fig. 1 Map depicting the distribution of Sr isotope compositions in basalts ($\text{SiO}_2 < 53.0$ wt% and $\text{MgO} > 4.0$ percent by weight) from the Western Great Basin. Of the samples with $\text{Zr/Ba} < 0.20$ (circles), 20 out of the 21 which occur within the area underlain by Proterozoic lithosphere have $^{87}\text{Sr}/^{86}\text{Sr} \geq 0.7060$ (closed circles), whereas all 13 samples occurring outside this region have $^{87}\text{Sr}/^{86}\text{Sr} < 0.7060$ (open circles). The isotopic compositions of samples with $\text{Zr/Ba} > 0.20$ (square symbols) do not appear to be controlled by the lithospheric boundary (open squares have $^{87}\text{Sr}/^{86}\text{Sr} < 0.7060$, closed squares have $^{87}\text{Sr}/^{86}\text{Sr} \geq 0.7060$). The location of the boundary of the North American craton is based on crustal isotopic studies^{19,38}, the observed distribution of Precambrian basement rocks³⁹, and the trace of the Roberts Mountains thrust²⁰. The broad bands numbered 1–6 correspond to the appropriately numbered stratigraphic columns of Fig. 3. The direction of Pacific relative to N. American present day plate motion is given by the solid arrow⁴⁰ (inset).

a peak at Nb and Ta (Fig. 2a). In this and other respects they are similar to late Cenozoic basalts from throughout the Basin and Range province for which an ocean-island basalt (OIB) type 'within-plate' source has been proposed¹². But most basalts from the WGB have trace-element characteristics that depart markedly from such a smooth pattern. In these samples the concentrations of the high field strength (HFS) elements Ta, Nb, Zr, Hf and Ti are much lower than elements of similar incompatibility. In some areas an attempt has been made to explain such patterns in continental flood basalts by the contamination of OIB magmas with continental crust¹³. However, the *ne-normative* character, high Ni and Cr contents, and high $\text{Mg}/(\text{Mg} + \text{Fe})$ ratios of many lavas from the WGB imply that they at least are relatively unmodified mantle melts. Moreover, to reduce the concentration of Nb in an OIB magma from a

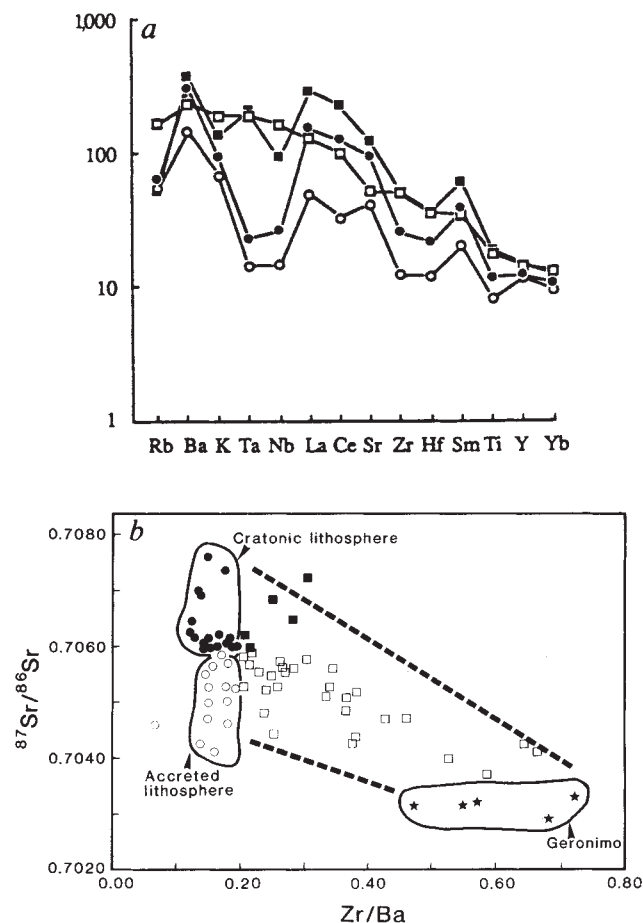


Fig. 2 a, Mantle normalized trace-element patterns of representative samples (Table 1) from the four groups distinguished in Fig. 1 using the same symbols: ■, L67-22; □, L82-139; ●, D85-15; ○, D85-67. Normalizing values from Menzies *et al.*³⁷. b, $^{87}\text{Sr}/^{86}\text{Sr}$ versus Zr/Ba ratio for WGB lavas. Data for the Geronimo Volcanic Field, southern Basin and Range, is taken from Kempton *et al.*⁴¹. Note that lavas from the WGB with $\text{Zr/Ba} \geq 0.20$ define a broad mixing array between a heterogeneous mantle lithosphere ($\text{Zr/Ba} < 0.20$) and Geronimo, OIB-type melts.

typical figure of 50 p.p.m. to the 10–20 p.p.m. range characteristic of this second group of basalts from the WGB, would require the addition of 150 to 400% crustal material, assuming the latter were entirely devoid of Nb. Thus the relative depletions in HFS elements are not a feature of crustal contamination processes, but rather of their mantle source regions.

Recent studies have shown that basalts generated at active continental margins may contain contributions from both subducted oceanic crust and the sub-continental lithosphere in addition to a component from the convecting upper mantle^{6,14}. However, the picture is further complicated because analyses of peridotite xenoliths^{15,16} and studies of continental flood basalts^{17,18} demonstrate that trace-element enriched lithospheric mantle may mimic some of the features of subduction related components (high large-ion lithophile to high-field strength element ratios), which testify to a role for previous subduction episodes in the development of mantle lithosphere. As the WGB spans the boundary between two crustal (and therefore presumably lithospheric mantle) provinces, it should be possible to determine whether such trace-element characteristics are due to contributions from subducted oceanic crust (either directly during the waning stages of slab subduction, or subsequently as

Table 1 Selected analyses of basalts from the Western Great Basin compared with an average for the Basin and Range Province

Sample number:	D85-6	D85-15	D85-16	L67-22	D85-40	D85-67	L82-82	L82-139	D85-152	D85-227	D85-228	D85-235	Basin and Range avg†
SiO ₂	47.77	47.88	46.51	48.50	52.42	49.79	52.26	52.97	50.16	50.36	47.08	48.98	46.63 (1.68)
TiO ₂	1.51	1.18	2.29	1.89	1.34	0.82	1.18	1.78	1.40	1.43	1.92	1.30	2.24 (0.35)
Al ₂ O ₃	16.18	15.27	13.96	16.60	16.13	16.51	18.33	18.06	16.37	17.06	16.33	15.79	15.68 (0.61)
Fe ₂ O ₃ *	10.95	9.59	12.13	11.80	7.74	9.42	9.43	8.12	9.83	8.40	9.47	8.90	11.67 (0.89)
MgO	8.32	9.86	8.63	5.93	6.14	9.65	5.12	4.46	7.29	7.05	9.37	9.14	8.02 (1.42)
CaO	9.84	11.20	10.65	8.49	9.44	9.04	8.15	7.32	8.37	8.27	9.37	10.76	9.06 (0.86)
Na ₂ O	3.37	2.67	2.89	3.65	3.65	2.66	3.57	4.25	3.56	3.76	3.86	2.89	3.62 (0.56)
K ₂ O	1.03	1.34	1.35	1.94	2.04	0.98	1.27	2.74	2.11	2.62	1.54	1.70	1.66 (0.5)
P ₂ O ₅	0.44	0.51	0.89	1.28	0.59	0.18	0.36	0.77	0.52	0.77	0.66	0.46	0.55 (0.12)
MnO	0.17	0.15	0.17	0.18	0.12	0.15	0.16	0.14	0.15	0.14	0.15	0.15	0.18 (0.01)
Total	00.58	99.65	99.47	100.26	99.61	99.20	99.83	100.61	99.76	99.86	99.75	100.07	99.31
Ni	110	153	150	68	57	198	25	29	107	113	192	157	138 (45)
Cr	234	449	310	90	124	533	54	59	214	188	285	331	208 (78)
Sr	616	1,099	1,237	1,438	1,155	482	1,031	602	849	1,053	1,006	1,006	666 (145)
Rb	17	22	28	18	32	19	17	58	49	45	20	26	28 (11)
Zr	159	172	247	349	194	82	126	340	203	239	246	176	230 (49)
Nb	15	9	26	32	11	5	8	56	14	20	17	12	52 (16)
Ba	570	1,168	1,240	1,416	1,284	555	1,854	898	1,338	1,306	725	1,063	457 (130)
Pb	4	7	5	15	10	12	12	8	9	10	10	5	8
Y	24	24	27	29	24	23	24	28	23	23	28	21	28 (3)
⁸⁷ Sr/ ⁸⁶ Sr	0.70563 ±4	0.707601 ±3	0.70601 ±3	0.70686 ±3	0.70614 ±3	0.70471 ±2	0.70460 ±3	0.70441 ±4	0.70528 ±2	0.70616 ±3	0.70534 ±3	0.70623 ±2	0.70302 (15)§
¹⁴³ Nd/ ¹⁴⁴ Nd	0.51247 ±1	0.51207 ±1	0.51246 ±1	0.51215 ±1		0.51270 ±1		0.51276 ±2	0.51253 ±3		0.51263 ±1	0.51250 ±2	0.51302 (1)§
Nepheline	3.02	2.85	2.81	1.22	0.00	0.00	0.00	0.00	1.66	3.48	7.84	2.96	6.06
Quartz	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orth	6.09	7.92	7.98	11.46	12.06	5.79	7.51	16.19	12.47	15.48	9.10	10.05	9.81
Plag	48.92	43.06	40.40	51.82	52.49	52.72	60.45	58.07	49.51	47.33	40.87	44.08	41.07
Diopside	16.26	21.42	21.08	8.72	17.36	10.89	6.46	7.52	12.77	11.41	15.71	20.41	16.04
Hypersthene	0.00	0.00	0.00	0.00	2.22	11.95	17.58	0.06	0.00	0.00	0.00	0.00	0.00
Olivine	18.55	18.49	17.64	17.41	9.56	13.42	2.31	11.49	16.92	15.47	18.58	16.72	17.27
Age (Myr)‡	4.9±0.6	9.5±0.4 (2)	10.5±0.4	<0.1 (1)	0.8±0.2	5–10 (1)	6.7±0.7	<0.1 (1)	6.8±0.3 (1)	4.8±0.6 (1)	1.3±0.2	2.6±0.1 (1)	
Long/lat	117° 14'/ 37° 50'	117° 52'/ 37° 27'	117° 55'/ 37° 27'	116° 37'/ 36° 44'	118° 07'/ 36° 59'	120° 29'/ 39° 42'	119° 12'/ 39° 13'	118° 49'/ 39° 33'	119° 13'/ 38° 42'	118° 02'/ 37° 44'	118° 11'/ 37° 04'	118° 52'/ 37° 34'	

* Total iron as Fe₂O₃

† Average and standard deviation (in parentheses) based on 97 samples from Lunar Craters (Nevada), the Mojave Desert (California), the Geronimo field (Arizona) and the Potrillo and other small fields in southern New Mexico (ref. 34).

‡ K-Ar whole rock analyses from (1) ref. 35, (2) ref. 36, and Table 1.

§ Sr and Nd isotopic averages and standard deviations based on samples from the South Basin & Range (ref. 12).

|| Menzies *et al.* (ref. 12).

relict material entrained within convecting upper mantle), or from the sub-continental lithospheric mantle. Unfortunately, the low concentrations of Nb in many WGB basalts make determinations of the abundance of this element susceptible to significant analytical errors. For this reason, the ratio between the HFS element Zr and the large-ion lithophile element Ba will be used to characterize the two end-member trace element geochemistries. Basalts with Zr/Ba < 0.20 show pronounced depletions in the HFS elements, which may reflect contributions either directly from subducted oceanic material, or from lithospheric mantle previously enriched above a subduction zone. Lavas with Zr/Ba ≥ 0.20 resemble OIB type magmas more, and are presumably derived from OIB type mantle, or they at least contain a component from such a source.

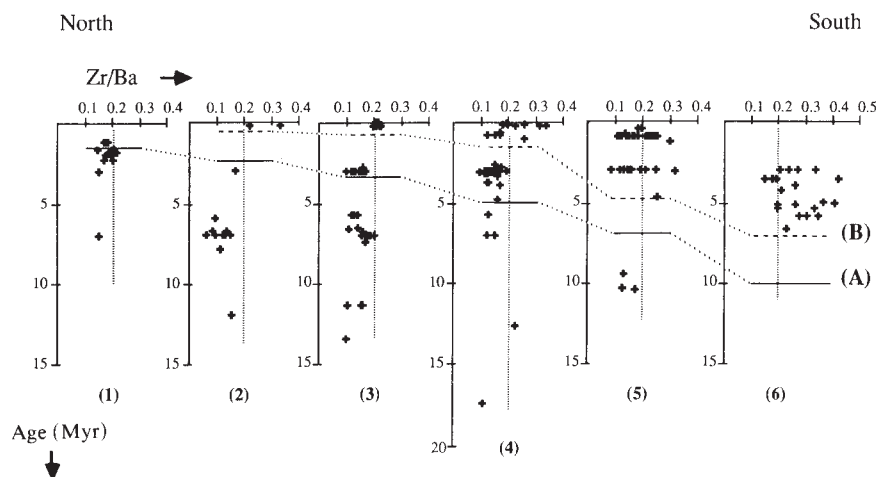
Figure 1 shows the distribution of Sr isotope compositions in WGB basalts in relation to the boundary between cratonic N. America (~1,800 Myr, ref. 19) and the allochthonous island-arc terranes accreted during the upper Palaeozoic²⁰. ⁸⁷Sr/⁸⁶Sr ratios in samples with Zr/Ba < 0.20 show a marked northwest-southeast geographic distribution: values ≥ 0.7060 occur only within the region underlain by Proterozoic crust and, with a single exception, values < 0.7060 occur outside this area. Likewise the western boundary of the N. American craton divides those basalts with ¹⁴³Nd/¹⁴⁴Nd > 0.51250 (to the north and west) from those with ¹⁴³Nd/¹⁴⁴Nd < 0.51250, which occur only on the craton. The high Sr contents (480–2,017 p.p.m.) and lack of any negative correlation between Sr and ⁸⁷Sr/⁸⁶Sr, in addition to the arguments against crustal contamination presented above

and elsewhere^{21,22}, suggest that the Sr isotope variations are sub-crustal in origin. The sympathetic lithospheric and Sr and Nd isotope provinciality supports earlier suggestions^{12,21–22} that these samples derived their isotopic characteristics from old mantle, likely to be in the sub-continental lithosphere. Alternative models, in which the ⁸⁷Sr/⁸⁶Sr in the WGB basalts reflects material from the subduction zone, would require the subduction of low ⁸⁷Sr/⁸⁶Sr material to the north and west of the cratonic margin and high ⁸⁷Sr/⁸⁶Sr material beneath the region to the east of this boundary, which is deemed unlikely. This distribution also argues against the possibility that a slab component was retained within the convecting asthenosphere.

Although some samples with Zr/Ba ≥ 0.20 also have elevated ⁸⁷Sr/⁸⁶Sr ratios, the majority are lower than 0.7060 and, significantly, they are similar, even where erupted through crust and lithosphere of different ages (Fig. 1). This implies derivation from a sub-lithospheric source within the convecting upper mantle. However, few of these basalts have Zr, Nb, TiO₂ as high, or Ba, Sr and ⁸⁷Sr/⁸⁶Sr as low as typical Basin and Range alkali basalts (see average analysis in Table 1). Rather, most of the samples with Zr/Ba ≥ 0.20 appear to represent mixtures between OIB type Basin and Range magmas and the overlying mantle lithosphere (Fig. 2b). The Zr/Ba = 0.20 'cut-off' is somewhat arbitrary, but useful in so far as it appears to distinguish those samples whose Sr isotope ratios reflect lithospheric provinciality from those in which that becomes swamped by a significant contribution from the underlying asthenosphere.

Within individual volcanic areas, the transition from older

Fig. 3 Series of 'stratigraphic' sections (age plotted against Zr/Ba ratio) along a line drawn normal to the direction of Pacific plate subduction. Each section represents $\sim 30'$ latitude and contains all the data from basaltic samples within the broad bands drawn on the map in Fig. 1 and numbered accordingly. Line A represents the projected time when the trailing edge of subducted lithosphere passed beneath each section^{23,42}. Line B marks the earliest occurrence of lavas with a predominantly within-plate geochemistry, as indicated by Zr/Ba ratios > 0.20 . K-Ar ages are taken from the compilation by Luedke and Smith³⁵, with additional analyses from Larsen³⁶ and Table 1. Full data tables for this and the preceding figures are available from the authors on request.



lithosphere-dominated ($Zr/Ba < 0.20$) to younger asthenosphere-dominated magmatism ($Zr/Ba \geq 0.20$) is related to the plate tectonic evolution of western North America. At around 28 Ma, the Pacific spreading ridge intersected the North American trench, off present-day Los Angeles. Subsequently, the Mendocino (trench-trench-transform) triple junction migrated northwards, replacing the destructive plate margin with the San Andreas transform system^{11,23}. The tectonic transition from a supra-subduction zone to within-plate environment can be correlated with the inferred position of the trailing edge of the descending slab (an oceanic transform boundary) beneath the WGB during the late Miocene-Pliocene. To evaluate the relationship between the change in tectonic setting and the onset of asthenospheric magmatism referred to above, a series of 'stratigraphic columns' were constructed along a NW-SE transect through the Western Great Basin (Fig. 3). Each stratigraphic column compares the ages of erupted lavas from a broad SE-NE oriented band across the WGB (as shown in Fig. 1) with their respective Zr/Ba ratios. Whereas lavas with a significant asthenospheric component ($Zr/Ba \geq 0.20$) were erupted at a relatively early stage in the south, they occur progressively later to the north, and have yet to appear in the most northerly section. Such a progression follows the northward migration of the trailing edge of subducted oceanic lithosphere (Fig. 3, line A), but only after a time-lag of 2–3 Myr. Figure 3 also shows that in several localities magmatism with a subduction-related geochemistry ($Zr/Ba < 0.20$) continued up to 5 Myr after slab subduction ceased.

The general correlation shown in Fig. 3 between the tectonic transition to a within-plate environment and the first eruption of within-plate alkali-basalts suggests that the subducted slab provides a physical barrier isolating the lithospheric mantle from magma source regions in the underlying asthenosphere. In one model, melts are present in the convecting upper mantle but they are only free to intrude and pass through the overlying lithosphere once the slab has been removed. But this fails to account for the 2–3 Myr lag period between the removal of the slab barrier and the first eruption of lavas with asthenospheric geochemistry because models of translithosphere magma ascent by induced fracture propagation suggest that an alkali basalt magma could rise to the surface from depths of ~ 100 km in a few weeks²⁴. An alternative model involves the formation and subsequent rise of asthenospheric diapirs. As subducted lithosphere moves away to the north beneath the WGB, the space it occupied is filled with upwelling asthenospheric material. A combination of decompression partial melting and induced convection give rise to density instabilities that develop into discrete mantle diapirs.

An earlier study of Basin and Range magmatism proposed that melt generation took place at relatively shallow depths,

between 40 and 60 km (ref. 25). This could suggest that the 2–3 Myr lag-period represents the time taken by upwelling asthenospheric diapirs in ascending from depths below the level of slab subduction into the zone of melt formation. For a relatively shallow angle of slab subduction $\sim 25^\circ$ (around 230 km depth beneath the WGB), this would imply ascent rates between 5 and 8 cm per annum. Quantitative estimates of the rate of asthenosphere upwelling are rare; although a value of 3 cm per annum used for the upwelling velocity of the mantle plume beneath Hawaii in a recent model²⁶ successfully accounted for the observed volume of volcanic products. In their model for the development of a mantle diapir beneath a spreading centre, Rabinowicz *et al.*²⁷ used a relative velocity of 10 cm per annum for the ascending diapir. The similarity between such estimates and the implied ascent rates for WGB asthenospheric diapirs is reassuring in the context of the model outlined above. Elsewhere it has been suggested²⁸ that OIB magmas are derived from diapirs that develop out of 'megalth' material at the 650 km seismic discontinuity. Such diapirs would be required to ascend at unreasonably high rates of between 15 and 24 cm per annum to accomplish such a journey within 2–3 Myr.

If our model relating the development of within-plate volcanism to the formation and subsequent ascent of diapirs from the level of the trailing edge of the subducted slab is essentially correct, then it carries an interesting implication for the earlier phase of lithospheric melt generation. Mantle xenolith data have shown that some parts of the sub-continental lithosphere are relatively depleted in major-element basaltic components compared with the oceanic upper mantle^{15,29,30}, and as such are thought to represent unlikely sites for large-scale melting³¹. However, we have shown that most of the volcanic activity within the WGB is characterized by a lithospheric trace-element signature and that some of these melts formed before the arrival of upwelling asthenospheric diapirs, which acted as a late external heat source. In agreement with a recent review³², it would appear that in some areas the sub-continental lithospheric mantle is sufficiently fertile to act as a source of both alkaline volcanism, as in the Western Great Basin, and large-scale flood tholeiite provinces, such as the Karoo¹⁸ and Parana³³.

We thank C. J. MacLeod for useful discussion and D. C. Rex for assistance with K-Ar analyses and G. A. Mahood for a thoughtful review. This work was undertaken during tenure of an Open University studentship (D.S.O.). Field work was partly supported by NSF. Isotope work at the Open University is supported by NERC.

Received 20 January; accepted 28 March 1988.

1. Pearce, J. A. & Norry, M. J. *Contr. Miner. Petrol.* **69**, 33–47 (1979).
2. Wood, D. A. *Earth planet. Sci. Lett.* **50**, 11–30 (1980).
3. Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, Berlin, 1981).

4. Pearce, J. A. in *Orogenic Andesites and Related Rocks* (ed. Thorpe, R. S.) 525–548 (Wiley, Chichester, 1982).
5. Fitton, J. G. & Dunlop, H. M. *Earth planet. Sci. Lett.* **72**, 23–38 (1985).
6. Pearce, J. A. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 230–249 (Shiva, Orpington, 1983).
7. Green, T. H. *J. Volcan. geotherm. Res.* **10**, 405–422 (1981).
8. Morris, J. D. & Hart, S. R. *Geochim. cosmochim. Acta* **47**, 2015–2030 (1983).
9. Van Kooten, G. K. *J. Petrol.* **21**, 651–684 (1980).
10. Van Kooten, G. K. *Contr. Miner. Petrol.* **76**, 378–385 (1981).
11. Atwater, T. *Bull. geol. Soc. Am.* **81**, 3513–3536 (1970).
12. Menzies, M. A., Leeman, W. P. & Hawkesworth, C. J. *Nature* **303**, 205–209 (1983).
13. Thompson, R. N. *et al.* in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 158–185 (Shiva, Orpington, 1983).
14. Ellam, R. M. & Hawkesworth, C. J. *Contr. Miner. Petrol.* (in the press).
15. Erlank, A. J. *et al.* in *Mantle Metasomatism* (eds Menzies, M. A. & Hawkesworth, C. J.) 221–311 (Academic, London, 1987).
16. Salters, V. F. M. & Shimizu, N. *Nature* (in the press).
17. Cox, K. G. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 139–157 (Shiva, Orpington, 1983).
18. Hawkesworth, C. J. *et al.* *Geol. Soc. S. Afr. Spec. Publ.* **13**, 341–354 (1984).
19. DePaolo, D. J. & Farmer, G. L. *Phil. Trans. R. Soc. A* **310**, 743–753 (1984).
20. Oldow, J. S. *Bull. geol. Soc. Am.* **95**, 174–185 (1984).
21. Hedge, C. E. & Noble, D. C. *Bull. geol. Soc. Am.* **82**, 3503–3510 (1971).
22. Leeman, W. P. *Bull. geol. Soc. Am.* **93**, 487–503 (1982).
23. Atwater, T. & Molnar, P. in *Conf. on Tectonic Problems of the San Andreas Fault System, Proc. Stanford Univ. Publ. Geol. Sci.* **13**, 136–148 (1973).
24. Spera, F. J. in *Mantle Metasomatism* (eds Menzies, M. A. & Hawkesworth, C. J.) 1–20 (Academic, London, 1987).
25. Leeman, W. P. & Rogers, J. J. W. *Contr. Miner. Petrol.* **25**, 1–24 (1970).
26. Ribe, N. M. & Smooke, M. D. *J. geophys. Res.* **92**, 6437–6443 (1987).
27. Rabinowicz, M., Ceuleneer, G. & Nicolas, A. *J. geophys. Res.* **92**, 3475–3486 (1987).
28. Ringwood, A. E. *J. Geol.* **90**, 611–643 (1982).
29. Nixon, P. H. & Boyd, F. R. in *Lesotho Kimberlites* (ed. Nixon, P. H.) 48–56 (Lesotho National Development Corp., Maseru, 1973).
30. Maaloe, S. & Aoki, K. *Contr. Miner. Petrol.* **63**, 161–173 (1979).
31. Jordan, T. H. *Nature* **274**, 544–548 (1978).
32. Hawkesworth, C. J. *et al.* *Nature* (submitted).
33. Hawkesworth, C. J. *et al.* *Nature* **322**, 356–359 (1986).
34. Fitton, J. G., James, D., Kempton, P. D. & Ormerod, D. S. *J. Petrology* (in the press).
35. Luedke, R. G. & Smith, R. L. *US Geol. Surv. Misc. Invest. Ser. Map* 1-1091-C (1981).
36. Larsen, N. W. thesis, Univ. Brigham Young (1979).
37. Menzies, M. A. *et al.* *Nature* **325**, 44–47 (1987).
38. Kistler, R. W. & Peterman, Z. E. *US Geol. Survey Prof. Paper* 1071, 17 (1978).
39. Bennett, V. C. & DePaolo, D. J. *Bull. geol. Soc. Am.* **99**, 674–685 (1987).
40. Minster, J. B., Jordan, T. H., Molnar, P. & Haines, E. *Geophys. J. R. astr. Soc.* **36**, 541–576 (1974).
41. Kempton, P. D., Dungan, M. A. & Blanchard, D. *Geol. Soc. Am. Spec. Pap.* **215**, 347–370 (1987).
42. Snyder, W. S., Dickinson, W. R. & Silberman, M. L. *Earth planet. Sci. Lett.* **32**, 91–106 (1976).

Westward drift, core motions and exchanges of angular momentum between core and mantle

D. Jault, C. Gire & J. L. Le Mouél

Laboratoire de Géomagnétisme, I P G P, 4 place Jussieu, 75232 Paris, France

The westward drift is one of the most well known features of the geomagnetic field. In this paper we come back to the apparent drift of the main field as seen at the Earth's surface, and show that there is no evidence for a body drift but rather clear evidence for a latitude-dependent drift. Computing the fluid flow at the core mantle boundary (CMB), in the frozen flux and geostrophic hypothesis, we also obtain, among other components, a clear differential zonal rotation of the fluid, symmetrical with respect to the Equator. This differential rotation changes with time, with a time constant of the order of 10 years. Dynamic considerations lead us to think that, during such a time interval, the change in the zonal toroidal flow inside the core consists of motions in which axial cylindrical annuli rotate rigidly about the Earth's axis of rotation. Then, for the 1969–85 time-span, the change in the core angular momentum is shown to balance the change in the mantle angular momentum.

Bullard *et al.*¹ stated that the westward drift u of the geomagnetic field presents no clear variation with latitude; this point has been reasserted^{2,3} since then. Bullard *et al.* compared two charts of the main field, one relative to 1907.5, the other one relative to 1945 and estimated the drift of the three elements X, Y, Z between 1907.5 and 1945, on nine different parallels.

Their results are illustrated by Fig. 1a, based on their Table 8. Whereas the curves relative to Y and Z (u_Y and u_Z) are similar (presenting higher values around the Equator), the one relative to X (u_X) displays two peaks, at the colatitudes 50° and 130°. Considering an average of u_X , u_Y , u_Z , Bullard and his co-authors inferred that no variation of u with latitude could be significantly determined.

In the present study the drift rate $u_E(\theta)$ of element E ($E = X, Y, Z$) for colatitude θ has been computed by minimizing the quantity

$$d^2 = \int_0^{2\pi} \left(\frac{\partial E}{\partial t} - u_E(\theta) \frac{\partial E}{\partial \phi} \right)^2 d\phi \quad (1)$$

We have used the models of Vestine *et al.*⁴ of the main and secular variation (SV) fields up to 1955, then IGRF models^{5–7}. The representative curves (Fig. 1b, c, d) look like the Bullard *et al.* curves (Fig. 1a). For each element, the general trend is quite coherent from one curve to another, and suggests a differential rotation. We can make a guess about the different behaviour of X on one hand, Y and Z on the other hand. A simple differential rotation $u(\theta)$ of the geomagnetic potential about the polar axis would indeed imply

$$\frac{\partial V}{\partial t} - u \frac{\partial V}{\partial \phi} = \frac{\partial Y}{\partial t} - u \frac{\partial Y}{\partial \phi} = \frac{\partial Z}{\partial t} - u \frac{\partial Z}{\partial \phi} = 0$$

but

$$\frac{\partial X}{\partial t} - u \frac{\partial X}{\partial \phi} = \frac{1}{a} \frac{\partial u}{\partial \theta} \frac{\partial V}{\partial \phi}$$

where a is the Earth's radius and r, θ, ϕ the usual spherical coordinates. Then if the latitude distribution of u were represented by a square window function, then u_X would display two peaks at the edges of the window, as observed in Fig. 1.

The body drift of the main field, as seen by Bullard *et al.*, was tentatively explained by a body rotation of the fluid of the core relative to the mantle at the CMB ($r = b$) (a true motion or a wave; see refs 1 and 8). Now from the observation of a latitude-dependent drift of the field, we rather expect a rotation of the fluid of the core at the CMB varying with latitude. Therefore we shall consider the zonal part of the toroidal ingredient of the flow computed at the CMB by inverting SV data in the frozen flux^{9,10} and geostrophic approximations^{11–13}. The horizontal flow velocity $\mathbf{v}$ at the CMB is written in the form

$$\mathbf{v} = \mathbf{t} + \mathbf{s} = \mathbf{n} \wedge \nabla T + \nabla S$$

T and S are the toroidal and consoidal scalars, $\mathbf{n}$ is the unit outward radial vector. In our computations the expansions of S and T in spherical harmonics are limited to degree 10. We have computed the flow at the CMB from 1969 to 1987, using the SV models listed in Table 1^{6,14–23}. We have then extracted from $\mathbf{v}$ its toroidal axisymmetric part $\mathbf{t}_{ax}$

$$\begin{aligned} \mathbf{t}_{ax}(\theta) &= -\mathbf{n} \wedge \nabla \left(\sum_{n=1}^{\infty} d_n P_n(\cos \theta) \right) \\ &= \sum_{n=1}^{\infty} \Theta_n^0 = t_{ax}(\theta) \hat{\phi} \end{aligned} \quad (2)$$

P_n is the Legendre polynomial of degree n , d_n are the coefficients (dimension = length²/time) of the expansion of the zonal part of T , $\hat{\phi}$ the unit azimuthal vector.

Figure 2 illustrates the distribution versus latitude of the angular velocity $w(\theta)$ ($w = t_{ax}/b \sin \theta$). $w(\theta)$ appears to be quite symmetrical about the Equator, larger at low latitudes (so are the apparent drift rates of the elements Y and Z). Its distribution may be represented by a rectangular window centred on the Equator (this distribution accounting for the two peaks noticed on the distribution of $u_X(\theta)$). The amplitude of the so-obtained zonal core motions is weaker (by at least a factor two) than the

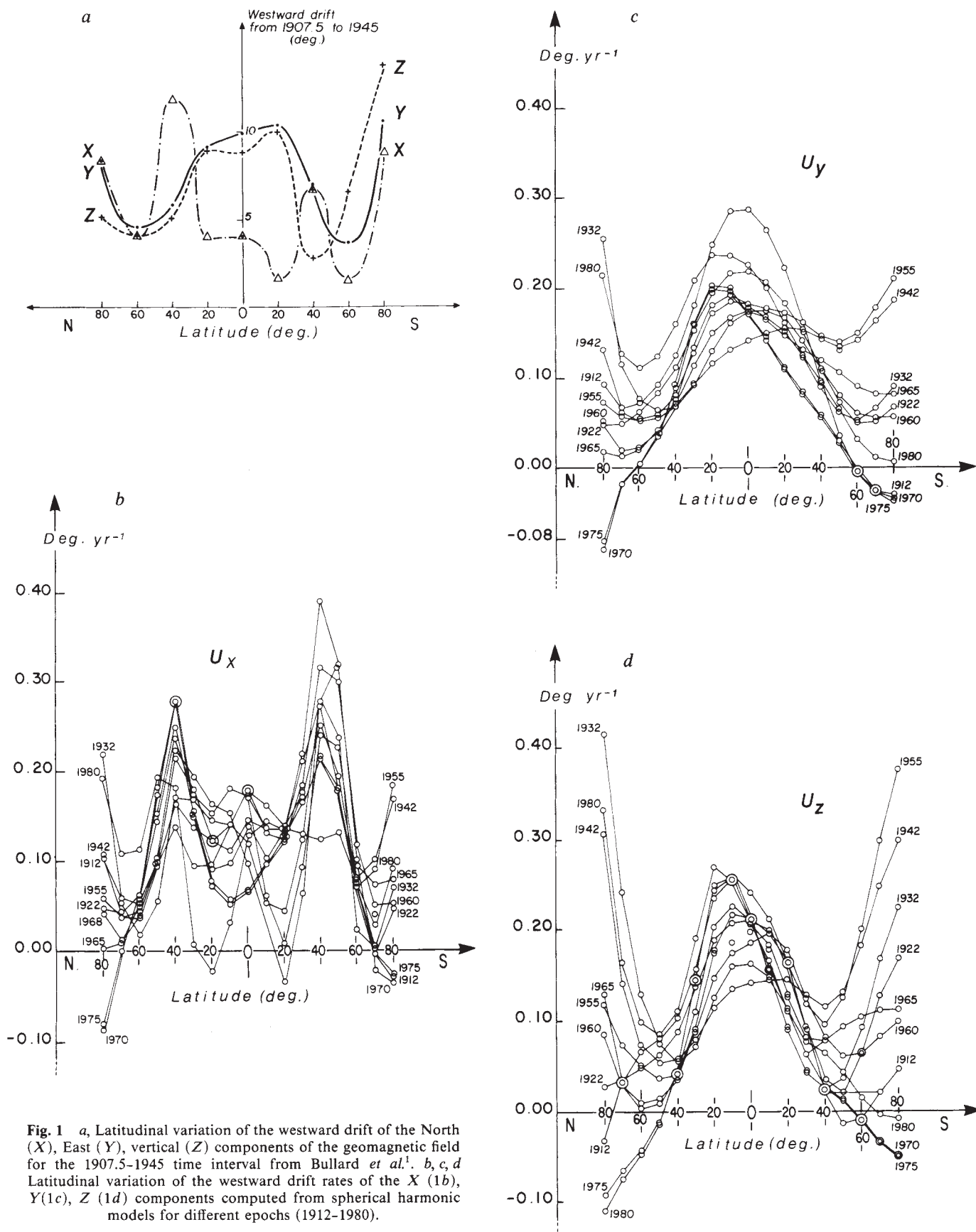


Fig. 1 *a*, Latitudinal variation of the westward drift of the North (X), East (Y), vertical (Z) components of the geomagnetic field for the 1907.5–1945 time interval from Bullard *et al.*¹. *b*, *c*, *d* Latitudinal variation of the westward drift rates of the X (*b*), Y (*c*), Z (*d*) components computed from spherical harmonic models for different epochs (1912–1980).

Table 1 Values of the angular rotations $d1/b^2$, $d3/b^2$, W_e (b being the core's radius) in degrees per year computed from different models, which are referenced

Year	Model	$d1/b^2$	$d3/b^2$	W_e	Present on Fig. 2?	Ref.
1969	SVHO 69	-0.073	0.014	-0.049	y	13
1970	SVAW 70	-0.074	0.015	-0.048	n	6
1970	SVDK 70	-0.064	0.011	-0.045	n	14
1975	SVDK 75	-0.069	0.009	-0.052	y	14
1980	GSFC 80	-0.089	0.018	-0.057	y	16
1980	USGS 80	-0.094	0.015	-0.068	y	15
1982	SVDK 82	-0.098	0.018	-0.067	n	14
1982	USGS 82	-0.117	0.027	-0.072	n	18
1982	SVIZM 82	-0.098	0.018	-0.067	n	19
1982	SVQKB 82	-0.128	0.027	-0.082	n	20
1985	IGRF 85	-0.116	0.030	-0.064	y	17
1987	WC 87	-0.16	0.047	-0.079	n	21
1987	SVBK 87	-0.159	0.048	-0.077	n	22

Symbol y means that the model has been used to draw Fig. 2, symbol n that it has not.

estimates of the apparent drift rate at the Earth's surface (ref 1); this point is reminiscent of the results of Kahle *et al.*²⁴. Indeed most of the secular variation must be attributed to non-zonal motions, whereas the estimates of the drift given by (1) try to represent the SV only by zonal motions.

Is it possible to infer some knowledge of the flow inside the core from the distribution of the zonal toroidal component t_{ax} of the so-computed surface flow at the CMB? In an incompressible fluid enclosed within a sphere, the system of equations governing the slow motions ($|\partial \mathbf{u}/\partial t| \ll |\mathbf{2}\Omega \times \mathbf{u}|$) is

$$2\rho\Omega \times \mathbf{u} = \mathbf{F} - \nabla p, \quad \text{div } \mathbf{u} = 0 \quad (3)$$

(where $\mathbf{u}$ is the flow in the frame rotating at angular velocity Ω , $\mathbf{F}$ the sum of all the forces, except the pressure contribution, applied to each element of fluid). Taylor²⁵ showed that the system (3) has a solution $\mathbf{u}$, $\mathbf{F}$ being given, if and only if

$$\Gamma = \mathbf{k} \cdot \int_{s=s_0} \mathbf{r} \times \mathbf{F} dS = 0 \quad (4)$$

(s is here a cylindrical radius, $\mathbf{k}$ the unit vector of the axis of rotation). In other terms, there is no mean torque Γ on any co-axial annular cylinder. Rigid rotations t_s of axial annular cylinders, about the axis of rotation appear as arbitrary motions (integration constant of the system (3)) (see ref 26). When this condition is not fulfilled ($\Gamma \neq 0$), it becomes necessary to introduce a relative acceleration. One can show, as Braginsky did²⁷, that only the relative accelerations t_s of the cylindrical annuli have to be taken into account. These accelerations respond to the average torque. Coriolis acceleration responds to all other forces through the system (3) slightly modified. In fact if $\Gamma \neq 0$, we can introduce $\mathbf{F}_0$ such that

$$\mathbf{F}_0 = F_0 \hat{\phi}, \quad \frac{\partial F_0}{\partial z} = 0, \quad \Gamma = \int_{s=s_0} (\mathbf{r} \times \mathbf{F}_0) \cdot \mathbf{k} dS$$

and the system (3) becomes

$$2\rho\Omega \times \mathbf{u} = \mathbf{F} - \mathbf{F}_0 - \nabla p, \quad \text{div } \mathbf{u} = 0, \quad \rho \partial t_s / \partial t = F_0$$

Roberts²⁸ conjectured that if this kind of flow is responsible for the westward drift, then the changes in the drift will be quicker in low-latitude zones (because of the reduced inertia of the external cylinders). The surface flow (at the CMB) corresponding to these cylindrical annuli motions is an axisymmetric toroidal flow; we will identify it with the motion t_{ax} computed above (implying that t_{ax} must be symmetrical about the Equator). Therefore, knowing the motion t_{ax} and its time variation at the

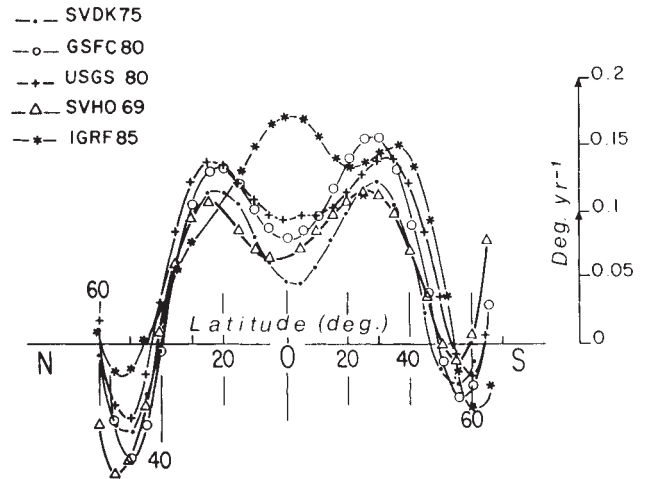


Fig. 2 Latitudinal variation of the zonal flow at the CMB obtained by inverting secular variation models for epochs 1970, 1975, 1980 and 1985 (see text).

CMB will allow us to compute the time changes in the cylindrical annuli motions and then the time changes in the angular momentum of the core.

The so-called decade variations in length of day are attributed for the main part to exchanges of angular momentum between the liquid core and the solid mantle^{1,28-32}. Nevertheless, atmospheric phenomena, especially South Pacific oscillations³³ and secular decrease of the rotation's rate³⁴ also contribute to the changes in angular momentum at the periods considered here. To estimate the changes $\Delta\sigma_m$ in the mantle angular momentum that must be balanced by changes $\Delta\sigma_c$ in the core angular momentum, this contribution has been removed. We have therefore used the estimates of the Earth's rotation rate for the 1969-87 period, as proposed by Feissel and Gavoret³⁵ who have separated the effect of pacific oscillations from the Earth's rotation data. The secular drift has been approximated by a linear trend (2 ms per century). The estimate of the angular momentum of the core depends only on the zonal toroidal component of the flow that, as shown above, is organized in co-axial annular cylinders whose surface expressions are known (t_{ax}). Then, if in a first approximation the core density is supposed to be uniform, the change in the core angular momentum depends only on the changes Δd_1 , Δd_3 of the two coefficients d_1 , d_3 of expansion (2)

$$\begin{aligned} \Delta\sigma_c &= -(8\pi/15)\rho b^3(\Delta d_1 + (12/7)\Delta d_3) \\ &= -(8\pi/15)\rho b^5\Delta w_e \end{aligned}$$

Δw_e is the change of the 'equivalent rotation' w_e

$$b^2 w_e = d_1 + (12/7)d_3 \quad (5)$$

To estimate w_e , we have used the estimates of d_1 and d_3 (Table 1) obtained when assuming in addition the symmetry of the zonal toroidal part of the motion. Fig. 3 displays the changes $\Delta\sigma_m$ and $\Delta\sigma_c$ for the time-span 1969-1985. Taking into account the uncertainties on both estimates, the relation $\Delta\sigma_m = -\Delta\sigma_c$ can be considered as reasonably checked. Overestimates of the westward drift rate changes had led to the idea that only a superficial shell, the thickness of which being taken as an adjustable parameter, would be involved in the balance^{1,31}. On the other hand, the flow considered here allows us to verify the balance between the angular momentum of the core and the mantle without the need of any variable parameter. We must investigate whether this relation holds for earlier epochs. In the present state of our computations the curves $t_{ax}(\theta)$ present an

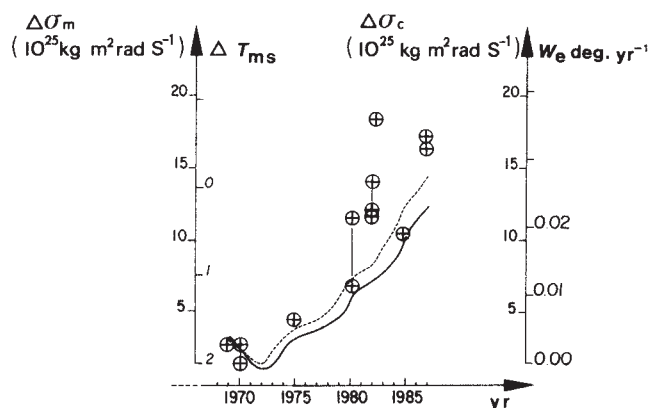


Fig. 3 Changes in the angular momentum of the mantle and of the core for the recent epoch (1969–1987). Solid line, $\Delta\sigma_m$ (see text) or ΔT (in ms); broken line, $\Delta\sigma_m$ (or ΔT) corrected from the secular decrease in length of day T ; $\oplus$, estimates of $-\Delta\sigma_c$.

increasing dissymmetry from 1965 to 1940 forbidding the strict application of formula (3); it could be because of the poor quality of the corresponding SV models (in particular the earlier models are based on very few observatories in the South Hemisphere (see Fig. 2 of Langel *et al.*¹⁷)). Nevertheless, the estimates of $\Delta\sigma_c$ before 1970 remain fairly well correlated with the estimates of $\Delta\sigma_m$; in particular we observe a decreasing trend of $(-\Delta\sigma_c)$ before 1970. Decade changes in the mantle angular momentum as observed during the past 20 yr can be balanced by a flow made of cylindrical axial annuli and compatible with the values computed at the CMB from SV data. The phenomenon responsible for the (differential) westward drift of the geomagnetic field would then be a genuine motion of fluid and not a MHD wave propagating westwards. In this paper we have only described the global budget of angular momentum; it is now necessary to study the exchange and coupling mechanisms both inside the core²⁷ and between the core and the mantle³⁰.

We thank G. Backus and R. Hide for useful discussions.

Received 15 October 1987; accepted 28 March 1988.

- Bullard, E. C., Friedman C., Gellman, H. & Nixon, J. *Phil. Trans. R. Soc. A* **243**, 67–92 (1950).
- Yukutake, T. *Bull. Earthquake Res. Inst.* **40**, 1–65 (1962).
- Yukutake, T. *Phys. Earth planet. Interiors* **24**, 253–258 (1981).
- Vestine, E. H., Laporte, L., Cooper, C., Lange, I. & Hendrix, W. C. *Carnegie Institution of Washington Publication* 578 (1947).
- McDonald, K. L. & Gunst, R. H. *ESA technical report*, 87 pp (1967).
- Hurwitz, L., Fabiano, E. B. & Peddie, N. W. *J. geophys. Res.* **79**, 1716–1717 (1974).
- Peddie, N. W. *J. Geomagn. Geoelect.*, Kyoto **33**, 607–611 (1981).
- Hide, R. *Phil. Trans. R. Soc. A* **259**, 615–650 (1966).
- Roberts, P. H. & Scott, S. J. *Geomagn. Geoelect.*, Kyoto **17**, 137–151 (1965).
- Backus, G. *Phil. Trans. R. Soc. A* **263**, 239–266 (1968).
- Le Mouél, J. L. *Nature* **311**, 734–735 (1984).
- Le Mouél, J. L., Gire, C. & Madden, T. *Phys. Earth planet. Interiors* **39**, 270–287 (1985).
- Backus, G. & Le Mouél, J. L. *Geophys. J. R. astr. Soc.* **85**, 617–628 (1986).
- Hodder, B. *Geophys. J. R. astr. Soc.* **65**, 763–776 (1981).
- Langel, R. A., Kerridge, D. J., Barraclough, D. R. & Malin S. R. *J. Geomagn. Geoelect.*, Kyoto **38**, 573–597 (1986).
- Peddie, N. & Fabiano, E. *J. Geomagn. Geoelect.*, Kyoto **34**, 357–364 (1982).
- Langel, R., Estes, R. & Mead, G. *J. Geomagn. Geoelect.*, Kyoto **34**, 327–349 (1982).
- IAGA Division I Working Group, *Geophys. J. R. astr. Soc.* **85**, 217–220 (1985).
- Peddie, N. W. & Zunde, A. K., *Phys. Earth planet. Interiors* **48**, 324–329, (1987).
- Golokov, V. P. & Kolomitseva, G. I. *Phys. Earth planet. Interiors* **48**, 320–323 (1987).
- Quinn, J. M., Kerridge, D. J. & Barraclough, D. R., *Phys. Earth planet. Interiors* **48**, 313–319 (1987).
- Quinn, J. M., Kerridge, D. J. & Barraclough, D. R. *Geophys. J. R. astr. Soc.* **87**, 1143–1157 (1987).
- Barraclough, D. R. & Kerridge, D. J. *Phys. Earth planet. Interiors* **48**, 306–312 (1986).
- Kahle, A. B., Vestine, E. H. & Ball, R. H. *J. geophys. Res.* **72**, 1095–1108 (1967).
- Taylor, J. B. *Proc. R. Soc. A* **274**, 274–283 (1963).
- Le Mouél, J. L., Jault, D. & Gire, C. *C. R. Acad. Sci., Paris* (II) **305**, 619–624, (1987).
- Braginsky, S. I. *Geomag. Aeron.* **10**, 3–12 (Engl. Trans pp (1–8)), (1970).
- Roberts, P. H. *J. Geomagn. Geoelect.*, Kyoto **24**, 231–259, 1972.
- Rochester, M. G. *Phil. Trans. R. Soc. A* **252**, 531–555, 1960.
- Hide, R. *Phil. Trans. R. Soc. Lond.* **A284**, 547–554 (1977).
- Runcorn, K. *Phil. Trans. R. Soc. Lond.* **A306**, 261–270 (1982).
- Stix, M. & Roberts, P. H. *Phys. Earth planet. Interiors* **36**, 49–60 (1984).
- Rosen, R. D. *et al. Science*, **225**, 411–414 (1984).
- Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge University Press, New York, 1980).
- Feissel, M. & Gavoret, J. *JUGG abstract*, Vancouver, Aug. 1987.

Genetic determination of guarding and undertaking in honey-bee colonies

Gene E. Robinson & Robert E. Page Jr

Department of Entomology, The Ohio State University,
1735 Neil Avenue, Columbus, Ohio 43210-1220, USA

Darwin¹ considered the dramatic differences in morphology and behaviour among sterile workers, the basis of colony division of labour in the social insects, to be a greater challenge to his theory of natural selection than the occurrence of worker sterility itself. Darwin's model for the evolution of these worker traits required: (1) heritable variation among workers within colonies; (2) variation in reproductive success among colonies due to different distributions of worker traits; and (3) changes in the distribution of worker traits within colonies due to colony-level selection. The role of genetics in this evolutionary process, unknown to Darwin, has still received little attention^{2,3}. Calderone and Page⁴ recently demonstrated differences in the pollen-collecting behaviour of honey bees from two artificially selected strains⁵ co-fostered in wild-type colonies to be a consequence of genotypic differences between workers. These differences were caused by an artificial selection process analogous to that proposed by Darwin. Their study established a foundation for understanding genetic mechanisms underlying the evolution of division of labour but did not demonstrate a genetic basis for division of labour between related members of colonies, the essential element of the darwinian model. Here we report previously undescribed genetic differences in task specialization between related members of *Apis mellifera* colonies. These results, which support the first requirement of the darwinian model for the evolution of colony organisation, suggest that the genetic structure of an insect society plays a fundamental, and previously unrecognized, role in the division of labour.

Behavioural differentiation among worker honey bees is in part age-based, but there is also variation among workers in the rate of behavioural development and the degree of task specialization at a particular age^{6,7}. Due to polyandry^{8–10} and sperm mixing^{10,11}, honey-bee colonies are assemblages of subfamilies, each subfamily consisting of the offspring of the queen and one of her mates. We identified the subfamily membership of individuals performing specific tasks, with allozyme analysis^{12,13}. Instrumental insemination¹⁴ of queen honey bees was used to establish experimental colonies with electrophoretically distinguishable subfamilies.

Two types of task specialists were collected ($n = 40$): workers guarding the nest entrance, and workers removing corpses. Both acts are readily identifiable according to simple, unambiguous criteria and are performed by a small subset of bees approximately the same age^{15,16}, enabling direct comparisons of guards and undertakers. Use of allozyme markers allowed blind behavioural observations.

Samples representing the proportion of adult workers in the colony from each subfamily ('controls') were obtained immediately after collecting task specialists. Individuals ($n = 40$) were collected randomly with respect to behaviour throughout the periphery of the hive, where bees of a similar age to guards and undertakers are normally located¹⁷. Age-matched controls ensured an appropriate reference group for guards and undertakers in the event of short-term fluctuations in subfamily frequencies^{10,11}. Colonies were sampled twice for task specialists and controls, at 1–2 week intervals.

Our results demonstrate that worker genotype influences the probability of guarding and removing corpses. Significant differences in the genetic composition of samples of guards and control bees were detected 9 out of 10 times; 4 out of 10 times in samples of undertakers and control bees; and 5 out of 10 times in samples of guards and undertakers (Fig. 1). Binomial

Fig. 1 Genotypic composition of samples of guards, undertakers and control workers from honey-bee colonies composed of electrophoretically distinguishable subfamilies. Statistical analyses: $\square$ = controls versus guards; $\blacksquare$ = guards versus undertakers; $\boxtimes$ = controls versus undertakers; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (based on G -tests of actual frequencies). In most cases $n = 40$, except: $n = 39$ for Colony 4442 guards and undertakers, trial 1, Colony 4438 guards and undertakers, trials 1 and 2, Colony 4456 control, trial 2; $n = 38$ for Colony 4439 undertakers, trial 1, Colony 4456 undertakers, trial 1; and $n = 36$ for Colony 4456 guards, trial 1. Significant inter-trial differences for guards: Colony 4438, $P < 0.02$; Colony 4439, $P < 0.01$; Colony 4456, $P < 0.05$; undertakers: Colony 4439, $P < 0.02$.

Methods. To establish experimental colonies, an electrophoretic survey of colonies in our research apiaries was conducted; eight unrelated queens were selected as mothers of haploid drones to be used for instrumental inseminations, and three additional unrelated queens were chosen as mothers of five virgin queens. Each virgin queen was inseminated with the semen of three unrelated drones, each drone bearing a different allozyme of malate dehydrogenase, 'slow' (S), 'medium' (M), and 'fast' (F)¹². Semen from each drone trio was pooled, diluted, and homogenized before insemination to stabilize the relative frequencies of different subfamilies over time^{35,36}. Two out of five inseminated queens (Colonies 4438 and 4456) were SS homozygotes and produced worker progeny with three distinct allozyme phenotypes: SS, SM and SF. The other three colonies had SF queens that produced three subfamilies with five allozyme phenotypes: SS, SM, SF, MF and FF. Workers sampled from these colonies were grouped into three subfamilies as follows: SM and MF bees were assigned to the M subfamily because the M allele could only be paternally inherited. SF bees were assigned to the S and F subfamilies in direct proportion to the ratio of SS to FF bees in each sample, because the maternal S and F alleles segregate in a 1:1 ratio. G -tests with all five allozyme phenotype classes gave identical results to those presented.

tests indicate these differences occurred too frequently to be explained by sampling error alone ($P < 0.01$, undertakers versus controls; $P < 0.001$, guards versus controls and guards versus undertakers). No allozyme marker was consistently associated with low or high levels of guarding or undertaking activity (Fig. 1).

Subfamily frequencies in the control samples of trials 1 and 2 did not change significantly ($P > 0.05$) for any colony (Fig. 1), suggesting that the genotypic composition of colonies was stable over the study period. Thus our results cannot be a consequence of possible age differences between control bees and task specialists coupled with age polyethism.

Inter-trial differences in the genotypic composition of guard and undertaker samples were observed in some cases (Fig. 1), suggesting environmental effects on the probability of task performance. A strong genotypic component to the observed subfamily variation is demonstrated by the persistence of specific subfamily biases for tasks from trial to trial. This was quantified by calculating the relative likelihood (RL) that a bee observed guarding or undertaking belongs to a given subfamily (i) as follows:

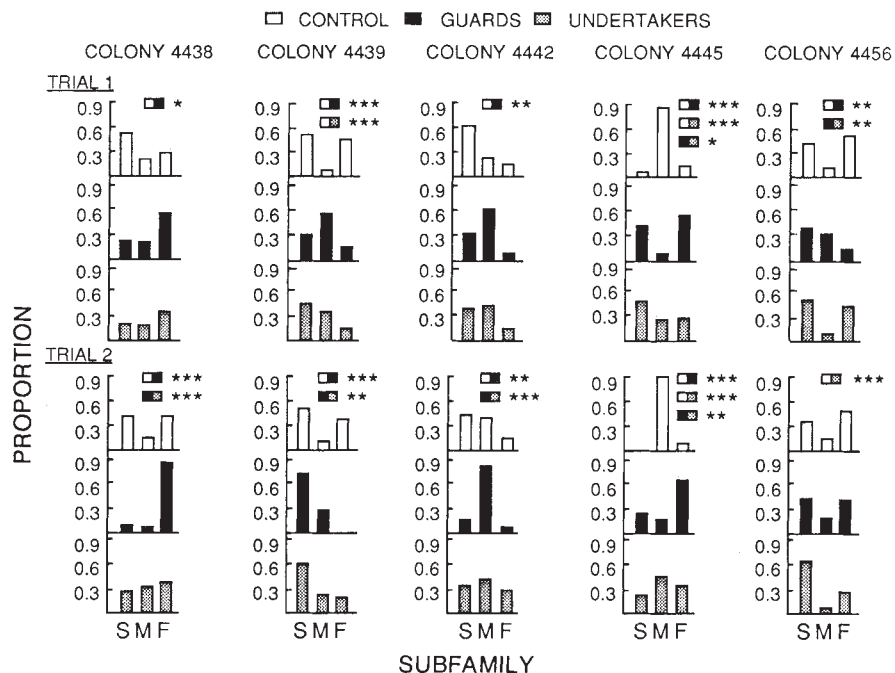
$$RL_i = \frac{L_i}{\sum_{j=1}^n L_j}$$

where

$$L = \frac{\text{proportion of subfamily } i \text{ or } j \text{ in task specialist sample}}{\text{proportion of subfamily } i \text{ or } j \text{ in control sample}}$$

n = number of subfamilies in a colony

Values of RL (two per colony for each task) were analysed with an analysis of variance (ANOVA); the proportion of the variance in RL attributable to subfamily membership was 89% and 81%, for guarding and undertaking, respectively. Additional evidence for highly constant subfamily biases in corpse removal is presented in Fig. 2.



There are at least two possible explanations for our results. First, genetic differences in subfamily spatial distribution within the nest may affect the probability of encountering stimuli associated with guarding and undertaking behaviour. Possible reasons for non-random spatial distribution independent of task performance include genotypically determined temperature preferen-

There are at least two possible explanations for our results. First, genetic differences in subfamily spatial distribution within the nest may affect the probability of encountering stimuli associated with guarding and undertaking behaviour. Possible reasons for non-random spatial distribution independent of task performance include genotypically determined temperature preferen-

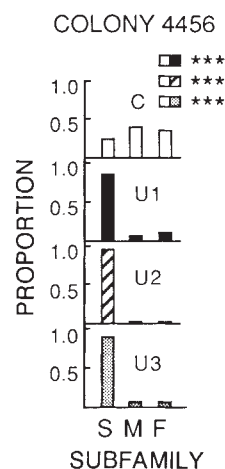


Fig. 2 Persistence of subfamily differences in task performance: genotypic composition of three samples of undertakers (U1-3) and a sample of control bees (C) collected four months after the first samples from this colony were taken (Fig. 1, Colony 4456). See Fig. 1 for explanation of statistical analyses. U1 and U2 were taken on 20 August 1987, U3 and C the next day. Bees from the S subfamily were over-represented in all undertaker samples relative to the control sample, in a pattern similar to that observed in the first trials with this colony (Fig. 1). The consistency of this effect, seen over the span of several worker lifetimes, despite differences in colony and environmental conditions associated with changing seasons, points to a strong genetic determinism for this task.

ces and kinship discrimination¹⁸. But in the only study on this question to date¹⁹, members of the same subfamily were not clumped within honey-bee nests.

Alternatively, we favour the idea that a colony's subfamilies differ in task performance because they have different distributions of behavioural response thresholds for the stimuli eliciting a task. This hypothesis is consistent with previous reports of genotypic²⁰ and hormonally mediated²¹ differences in honey-bee sensitivity to task-associated stimuli. Based on this possibility, we propose a mechanistic explanation for one of the most interesting and enigmatic forms of division of labour in the insect societies, the occurrence of individuals like guards and undertakers that perform jobs most colony members do not²²: these individuals have rare genotypes that result in low response thresholds.

The discovery that intra-colonial genetic variation influences the organisation of labour challenges the prevailing model of highly eusocial insect colonies as superorganisms responding to environmental stimuli with behaviourally totipotent individuals^{23,24}. For some tasks, particularly those involving large numbers of workers, the additive effects of all genotypes may contribute to the colony 'phenotype'. Non-additive behavioural dominance²⁵ effects of one or a few distinct genotypes may determine the colonial phenotype for other tasks. Negative feedback within colonies coupled with behavioural dominance may in particular affect the genetic structure of groups performing relatively rare jobs like guarding and undertaking, because the activities of a small number of bees may satisfy colony needs and maintain the stimulus level below the thresholds of less sensitive individuals. Feedback loops of task regulation may also result in less sensitive workers performing a particular task if a colony experiences a greater need for that task or lacks a subfamily with a high probability of performing it. Worker bees exhibit impressive behavioural plasticity⁷, but findings of genetic variation in colony behaviour^{5,26} suggest that workers do not converge on some average level of task performance in a colony, but rather express genetically determined differences. Colony behaviour is thus the product of the behaviour of its individuals, and differences between colonies are in part due to differences in the subfamily distribution of worker traits within colonies.

Our results support and extend previous findings of genetic variation in worker behaviour among different races and artificially selected strains of honey bees^{26,27}. Together with those of Calderone and Page⁴, they also demonstrate that genetic differences between workers are the 'grist' for the evolution of colony organisation. Heritable differences between colonies in the division of labour could also be the result of queen effects on workers²⁸⁻³⁰, but we know of no supporting evidence.

If, in the course of evolution, worker behaviour has been under strong colony-level selection, the large amount of intra-colonial genotypic variability in task performance we observed needs to be explained. Worker heterogeneity may be an artefact of polyandry, itself selected for by factors unrelated to the organisation of work^{31,32}. If this is true, it may be that colony-level selection does not reduce genetic variation because heritabilities are low due to behavioural dominance, or because fluctuations in the environment preclude the emergence of a single optimal genotype³³. Alternatively, increased intra-colonial genetic variability may have selective value, if a collection of specialized genotypes performs more efficiently under a range of environmental conditions than a single, more generalized genotype^{2,31,34}. The adaptive significance of our findings is not clear, but it is apparent that colony genetic structure affects the division of labour in at least one polyandrous insect society.

Since submitting this manuscript we have learned that Frumhoff and Baker³⁷ report similar results.

We thank M. K. Fondrk for technical assistance, J. A. Gianelos for assistance with electrophoretic analyses, N. W. Calderone, R. D. Mitchell and T. D. Seeley for helpful discussions, and

M. D. Breed, N. W. Calderone, R. Nowogrodzki, P. W. Sherman and P. K. Visscher for suggestions that improved the manuscript. This study was supported in part by an Ohio State University Postdoctoral Fellowship to G.E.R., and grants from the NSF and Ohio State University to R.E.P.

Received 14 January; accepted 5 April 1988.

1. Darwin, C. *The Origin of Species* 250-257 (Mentor, New York, 1958).
2. Crozier, R. H. & Consul, P. C. *Theor. pop. Biol.* **10**, 1-9 (1976).
3. Owen, R. E. *Theor. pop. Biol.* **29**, 198-234 (1986).
4. Calderone, N. W. & Page, R. E. *Behav. Ecol. Sociobiol.* **22**, 17-25 (1988).
5. Hellmich, R. L., Kulincevic, J. M. & Rothenbuhler, W. C. *J. Hered.* **76**, 155-158 (1985).
6. Seeley, T. D. *Honeybee Ecology: A Study of Adaptation in Social Life* 31-36 (Princeton University Press, Princeton, 1985).
7. Winston, M. L. *The Biology of the Honey Bee* 89-110 (Harvard University Press, Cambridge, 1987).
8. Taber, S. & Wendel, J. *J. econ. Ent.* **51**, 786-789 (1958).
9. Adams, J., Rothman, E. D., Kerr, W. E. & Paulino, Z. L. *Genetics* **86**, 563-596 (1977).
10. Page, R. E. & Metcalf, R. A. *Am. Nat.* **119**, 263-281 (1982).
11. Laidlaw, H. H. & Page, R. E. *Genetics* **108**, 985-997 (1984).
12. Contel, E. P. B., Mestriner, M. A. & Martins, E. *Biochem. Genet.* **15**, 859-876 (1977).
13. Romero-Severson, J., Hogg, D. B., Kingsley, P. C. & Schwalbe, C. P. *Ann. ent. Soc. Am.* **79**, 364-368 (1986).
14. Laidlaw, H. H. *Instrumental Insemination of Honey Bee Queens* (Dadant, Hamilton, 1977).
15. Visscher, P. K. *Anim. Behav.* **31**, 1070-1076 (1983).
16. Moore, A. J., Breed, M. D. & Moore, M. J. *Anim. Behav.* **35**, 1159-1167 (1987).
17. Seeley, T. D. *Behav. Ecol. Sociobiol.* **11**, 287-293 (1982).
18. Page, R. E. & Breed, M. D. *Trends Ecol. Evol.* **2**, 272-275 (1987).
19. Frumhoff, P. C. & Schneider, S. *Anim. Behav.* **35**, 255-262 (1987).
20. Collins, A. M. *J. Apicult. Res.* **18**, 285-291 (1979).
21. Robinson, G. E. *J. comp. Physiol. A* **160**, 613-619 (1987).
22. Oster, G. F. & Wilson, E. O. *Caste and Ecology in the Social Insects* 151-155 (Princeton University Press, Princeton, 1978).
23. Wilson, E. O. in *Experimental Behavioral Ecology and Sociobiology* (eds Hölldobler, B. & Lindauer, M.) 307-325 (Sinauer, Sunderland, 1985).
24. Wilson, E. O. *Science* **228**, 1489-1495 (1985).
25. Craig, R. *Am. Nat.* **116**, 311-323 (1980).
26. Rinderer, T. E. & Collins, A. M. in *Bee Genetics and Breeding* (ed. Rinderer, T. E.) 155-177 (Academic, New York, 1986).
27. Winston, M. L. & Katz, S. J. *Behav. Ecol. Sociobiol.* **10**, 125-129 (1982).
28. Free, J. B. *Anim. Behav.* **15**, 134-144 (1967).
29. Jaycox, E. R. *Ann. ent. Soc. Am.* **63**, 222-228 (1970).
30. Free, J. B., Ferguson, A. W. & Simpkins, J. R. *Physiol. Ent.* **9**, 387-394 (1984).
31. Crozier, R. H. & Page, R. E. *Behav. Ecol. Sociobiol.* **18**, 105-115 (1985).
32. Sherman, P. W., Seeley, T. D. & Reeve, H. K. *Am. Nat.* (in the press).
33. Hedrick, P. W. *A. Rev. Ecol. Syst.* **17**, 535-566 (1986).
34. Blum, M. S. *Proc. 8th int. Congr. IUSSI* 54-59 (Wageningen, Holland, 1977).
35. Kaftanoglu, O. & Peng, Y. S. *J. Apicult. Res.* **19**, 205-211 (1980).
36. Moritz, R. F. A. *J. Apicult. Res.* **24**, 249-255 (1983).
37. Frumhoff, P. C. & Baker, J. *Nature* **333**, 358-361 (1988).

A genetic component to division of labour within honey bee colonies

Peter C. Frumhoff*[‡] & Jayne Baker†[‡]

* Ecology Graduate Group and † Department of Entomology, University of California, Davis, California 95616, USA

Division of labour among nestmate workers is central to the colonial organization and ecological success of the eusocial Hymenoptera (ants, bees and wasps)¹. Workers characteristically divide labour through (1) ontogenetic changes in individual behaviour^{2,3} and (2) inter-individual variation in behavioural repertoire^{3,4}. On the basis of current evidence, optimization models of colony demography^{3,5,6} assume that variation among nestmates in behavioural repertoire arises solely through variation in environmental conditions, such as larval nutrition (inducing size-mediated behavioural differences in many ants)^{3,4,7} and adult experience (effecting behavioural differences among morphologically similar nestmates)⁸. A possible genetic component to division of labour, however, has received little study⁹. Yet, the degree of genetic heterogeneity among workers within Hymenopteran colonies is often extremely high¹⁰, a consequence of multiple mating by the queen (polyandry^{11,12}) and/or the presence of multiple laying

[‡] Present addresses: Departments of Organismic and Evolutionary Biology and Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA (P.C.F.); Department of Zoology NJ15, University of Washington, Seattle, Washington 98195, USA (J.B.).

queens (polygyny¹³). Here we report evidence of genetically based variation in task performance among nestmate workers in the polyandrous honey bee *Apis mellifera* L. Such variation may be an important component to division of labour within genetically heterogeneous colonies.

A honey bee queen typically mates with 7–17 males (drones) in a mating aggregation containing drones from several colonies¹⁴. Semen from mated drones mix in the queen's spermatheca, resulting in a colony composed of multiple worker patrilines^{11,15}. We observed worker behaviour in two colonies, each containing two patrilines, at the Bee Biology Facility of the University of California at Davis. We monitored workers performing two common tasks: social grooming, which may facilitate removal of debris and ectoparasites¹⁶, and trophallaxis (liquid food exchange), which results in the rapid distribution of nutrients and pheromones^{2,17}.

Each colony was established by artificially inseminating a virgin queen with semen from two drones, with queen and drones chosen so that half-sister worker nestmates would differ at the *cordovan* locus and express different cuticle colour patterns, termed black and cordovan (Fig. 1a). Each inseminated queen was then introduced into a queenless ten-frame field colony where she was allowed to oviposit for a period greater than worker life-span (>8 weeks) until all of the colony's workers ($n \sim 10,000$) were her daughters.

Workers between 2–30 days post-eclosion commonly groom and feed nestmates (refs 18, 19 and unpublished observations). To monitor the behaviour of individuals across this age-range, we collected callow workers (<24 h post-eclosion) for 18 (Colony 1) or 21 consecutive days (Colony 2) and tagged each on the dorsal thorax with a plastic disk imprinted with a unique number and colour combination (Fig. 1, legend). In each colony, the relative proportions of black and cordovan workers that emerged and were tagged varied significantly across days, presumably because sperm from different drones mix non-randomly within the queen's spermatheca²⁰ (Fig. 1, legend). For each age cohort, these proportions were assumed to remain constant throughout the study; the absolute number of tagged workers presumably decreased through mortality.

Immediately after returning the final group of tagged workers to their colony, we transferred each colony into a four-frame observation hive. Behavioural observations began 24–36 h following transfer. Pairs of workers engaging in social grooming or trophallaxis were located by a systematic scan of a 4 × 4 cm grid overlaid on the glass walls of each hive. To avoid repeated observations of any worker pair, the duration of each scan (10 min) exceeded the duration of grooming and feeding bouts (refs 21, 22 and unpublished observations). We recorded 1,649 interactions involving a tagged donor (a worker grooming or feeding nestmates) during 6–8 daylight hours per day over eight consecutive days in Colony 1 and seven consecutive days in Colony 2. The patriline and identity of all tagged donors and recipients were recorded and their ages determined by counting the number of days between eclosion and observation.

Because all of a colony's workers are reared in a common environment, any consistent patrilineal differences in task performance must have a genetic basis. For each task, we compared the observed numbers of black and cordovan donors of each age with the numbers expected if workers in both patrilines were equally likely to perform it (Fig. 2, legend). Observed and expected values were summed across ages to compare patrilines in three different age-classes: young workers that engaged in a variety of tasks within the colony (Age I, 2–9 days post-eclosion), workers in transition between within-colony work, nest guarding and foraging (Age II, 10–18 days) and older workers that were primarily foragers (Age III, 19–30 days).

Log-linear contingency table analyses²³ reveal striking patrilineal differences in the propensity of workers to groom nestmates (Fig. 2). In each colony, a significantly greater than expected proportion of groomers were in the black patriline

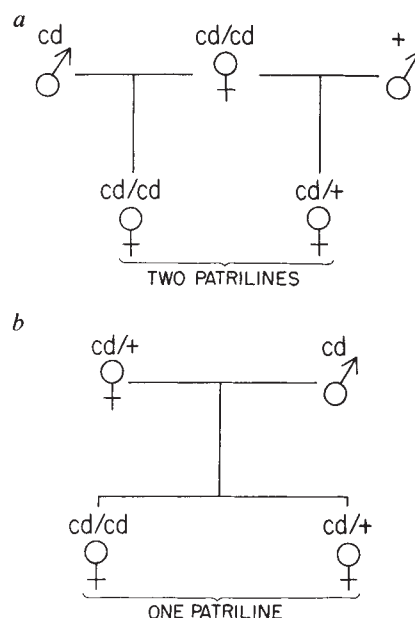


Fig. 1 Mating design, indicating genotype at the *cordovan* locus of each queen and drone. In homozygous or hemizygous (haploid) form, *cordovan* affects cuticle colour, producing a brown phenotype on typically black regions³⁸. **a**, Two-patriline colonies (Colonies 1 and 2) were produced from queens homozygous for the *cordovan* (*cd*) allele mated with one cordovan drone and one black (+) drone. **b**, One-patriline control colonies (Colonies 3 and 4) were produced from a heterozygous cordovan queen mated to one cordovan drone. For each mating, the queen and drone(s) came from apiaries separated by 5–15 km, within the range reported for natural matings¹⁷. Replicate queens and drones were from different outbred colonies within an apiary and of unknown relatedness to each other. Workers in Colonies 1 and 2 were tagged <24 h after emerging from a sealed brood frame placed in a 31 °C incubator (number of workers tagged in Colony 1: total = 3,816; $\bar{X}$ = 212.00 day⁻¹, s.d. = 25.65, range = 182–266 day⁻¹; Colony 2: total = 3,666; $\bar{X}$ = 174.57 day⁻¹, s.d. = 92.38, range = 17–323 day⁻¹). Relative proportions of emerging black and cordovan workers varied significantly across days in Colony 1 (mean proportion of black workers = 0.407; range = 0.286–0.542; G = 76.50, G = 76.50, P < 0.001) and Colony 2 (mean proportion of black workers = 0.390; range = 0.253–0.635; G = 168.17, P < 0.001). Relative proportions of emerging black and cordovan workers did not vary significantly across days in Colony 3 (mean proportion of black workers = 0.493; range = 0.450–0.573; G = 5.016, P > 0.10) nor deviate significantly from the expected 50:50 ratio (G = 0.177; P > 0.50). Emerging workers were not counted in Colony 4.

(Colony 1: G = 73.97, P < 0.001; Colony 2: G = 29.43, P < 0.001). Black workers comprised only 33.81% (Colony 1) and 38.59% (Colony 2) of the tagged worker population, yet comprised 67.30% (Colony 1) and 55.65% (Colony 2) of all groomers. Patrilineal differences were significantly greater in Colony 1 than Colony 2 (G = 7.26, P = 0.026) and did not vary significantly among workers of different age-classes (G = 0.31, P = 0.84).

We observed no consistent patrilineal differences in the propensity of workers to feed nestmates (Fig. 2). The relative numbers of black and cordovan workers observed feeding did not differ significantly from expected in any age-class of Colony 1 and differed only slightly and in varying direction from expected among younger (Age I and Age II) workers in Colony 2.

Patrilines in colonies with naturally mated queens may not typically carry different *cordovan* genotypes, raising the possibility that behavioural differences might arise simply as an artefact of atypical variation at or closely linked to the *cordovan* locus. To assess this, we constructed two control colonies (Colonies 3 and 4): each contained a queen artificially inseminated by a

single drone, with queen and drone chosen so that 50% of the daughter workers would be black and 50% cordovan (Fig. 1b). Thus, black and cordovan nestmates belonged to a single patriline; however, they carried the same genotypes at the *cordovan* locus as their counterparts in each two-patriline colony.

Workers in the one-patriline colonies were not tagged. In all other respects, one- and two-patriline colonies were treated equivalently and worker behaviour was monitored by observers unfamiliar with each colony's genetic composition. In Colonies 3 and 4, we recorded 2,682 bouts of social grooming and trophallaxis over 53 h. For both tasks in each colony, the number of interactions involving black and cordovan donors did not significantly differ from the expected 50:50 ratio (Grooming: $G = 0.224$, $P > 0.50$, $n = 541$ interactions (Colony 3); $G = 0.008$, $P > 0.50$, $n = 508$ (Colony 4). Trophallaxis: $G = 0.884$, $P > 0.50$, $n = 951$ (Colony 3); $G = 0.224$, $P > 0.50$, $n = 682$ (Colony 4). Thus, patrilineal differences in grooming behaviour were not an artefact of the cordovan marker.

Previously, studies of honey bees have revealed a genetic component to behavioural variation among workers of different geographic races^{24,25} and artificially selected strains²⁶⁻²⁸. The present results constitute the first evidence of genetically-based behavioural variation between nestmate worker patrilines. This variation is directly counter to the assumption of genetic equipotentiality underlying current optimization models of colony demography^{3,5,6}.

Patrilineal differences in task performance may simply be genetic 'noise', an unselected by-product of multiple mating favoured by any of several possible factors²⁹⁻³⁰. Thus, genetic heterogeneity may constitute an important constraint on the evolution of optimal task allocation. Alternatively, behavioural genetic variation among nestmates may enhance colony fitness³¹. If so, clues to the functional significance of such variation may lie in characteristic differences between tasks for which it is and is not strongly expressed.

One characteristic feature of social grooming is that a small proportion of workers within a patriline tend to specialize as groomers, repeatedly grooming nestmates over time. A comparison of the number of tagged individuals observed grooming nestmates 1, 2, ... n times reveals, for three of four patrilines,

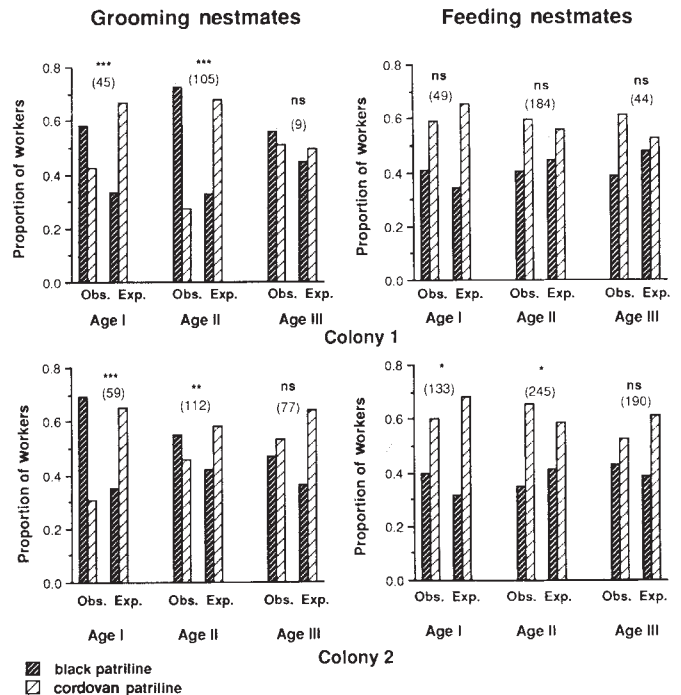


Fig. 2 Differences between worker patrilines in task performance. Observed proportions based on the number of different black versus cordovan tagged workers of each age-class observed grooming or feeding nestmates. Workers observed performing a task on more than one day were assigned to the day (and age) at which they were observed most frequently, with ties decided by random choice. Expected proportions calculated by multiplying the total number of tagged workers of each age observed performing a task by the known ratio of black to cordovan tagged workers of that age within the colony. Age I, 2–9 days post-eclosion; Age II, 10–18 days; Age III, 19–30 days. *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. P -values based on G -test for goodness of fit with the Williams' correction or on binomial probabilities when $N < 25$ (ref. 37). Sample sizes (total number of tagged workers observed) are in parentheses.

Table 1 Individual variation in social grooming and trophallaxis

Task	Colony	Patriline	Observed versus expected (Poisson) distribution	1	2	3	4	5	6+	P
Grooming	1	black	observed	76	15	4	2	2	8	<0.001
			expected	40.63	34.41	19.13	8.22	2.80	1.82	
			deviation from expected	+	—	—	—	—	+	
Grooming	1	cordovan	observed	49	3	0	0	0	0	>0.50
			expected	49.10	2.79	0.11	0.00	0.00	0.00	
			deviation from expected	—	+	—	—	—	+	
Grooming	2	black	observed	104	17	5	3	6	4	<0.001
			expected	69.63	43.22	17.89	5.55	1.38	0.34	
			deviation from expected	+	—	—	—	+	+	
Grooming	2	cordovan	observed	90	9	5	2	0	4	<0.001
			expected	62.02	32.68	11.48	3.03	0.64	0.15	
			deviation from expected	+	—	—	—	—	+	
Trophallaxis	1	black	observed	101	13	0	0	0	0	>0.50
			expected	101.93	11.21	0.82	0.05	0.00	0.00	
			deviation from expected	+	—	+	+	—	—	
Trophallaxis	1	cordovan	observed	145	18	0	0	0	0	>0.50
			expected	146.29	15.54	1.10	0.06	0	0	
			deviation from expected	—	+	—	—	—	—	
Trophallaxis	2	black	observed	202	17	2	0	0	0	>0.50
			expected	201.26	18.55	1.13	0.05	0	0	
			deviation from expected	—	+	+	—	—	—	
Trophallaxis	2	cordovan	observed	306	35	4	2	0	0	>0.50
			expected	302.24	40.77	3.67	0.25	0.07	0	
			deviation from expected	+	—	+	+	—	—	

Frequency distribution of number of different tagged individuals observed performing tasks 1, 2, ... n times compared to numbers expected if individuals do not differ in propensity to repeatedly perform task. Expected values based on a truncated Poisson distribution³⁶. P -values based upon Kolmogorov-Smirnov goodness of fit comparison of observed and expected distributions³⁷.

a significantly nonrandom distribution (Table 1). A greater than expected number of workers were observed grooming once ('nonspecialists') and several times ('specialists'), with fewer than expected observed an intermediate number of times. In Colony 2, black and cordovan nestmates observed grooming did not differ significantly in their propensity to specialize ($P > 0.10$, Mann-Whitney U test). Specialists did not differ significantly in age-class from non-specialists in either colony ($P > 0.50$, G -test); they comprised only 5.29% (Colony 1) and 5.96% (Colony 2) of all groomers, but engaged in 31.65% (Colony 1) and 29.05% (Colony 2) of all grooming bouts. In marked contrast, individual workers did not specialize on feeding nestmates (Table 1).

Social insect workers often specialize on tasks that appear to require the learning of particular skills for efficient performance^{3,4}. The coupling of patrilineal differences in the propensity to groom with individual specialization suggests a corollary to the proposal^{29,30} that the genetically heterogeneous worker population produced by a multiply mated queen may constitute a selective advantage to colonies in variable environments. Specifically, a colony's ability to cope with unpredictable events that require the reaction of a varying number of experienced workers (as may the response to ectoparasites by grooming specialists) may be enhanced by genetic variation in the threshold at which workers first respond to appropriate stimuli. Moreover, multiple mating by queens need not be a strategy to select for particular genotypes; genetic variation in initial response threshold rather than specific thresholds *per se* may suffice.

Determining what role behavioural genetic variation plays in colony social organization will ultimately require an assessment of the array of tasks for which it is maintained, the set of ecological³² and social^{33,34} contexts in which it is expressed, and the effect of such variation on patriline and colony fitness.

Since submitting this manuscript, we have learned that Robinson and Page³⁵ report similar results.

We thank S. Austad, W. Bossert, P. Calabi, N. Carlin, H. Dingle, R. Grosberg, D. Perlman, S. Schneider, J. Stamps, B. Waldman and P. Ward for discussion and comments.

Received 1 February; accepted 17 April 1988.

- Wilson, E. O. *Science* **228**, 1489–1495 (1985).
- Michener, C. D. *The Social Behavior of the Bees* (Harvard University Press, Cambridge, 1974).
- Oster, G. F. & Wilson, E. O. *Caste and Ecology in the Social Insects* (Princeton University Press, Princeton, 1978).
- Calabi, P. in *Advances in Myrmecology* (ed. Arnett, R. H. Jr) (Brill, Leiden, 1988).
- Wilson, E. O. *Fortschr. Zool.* **31**, 307–324 (1985).
- Franks, N. *Trends Ecol. Evol.* **2**, 72–75 (1987).
- Brian, M. V. in *Social Insects Vol. 1* (ed. Hermann, H. R.) 121–222 (Academic, New York, 1979).
- Jeanne, R. L. *Interindividual Behavioral Variability in Social Insects* (Westview, Boulder, 1988).
- Craig, R. & Crozier, R. H. *Isozyme Bull.* **11**, 66–67 (1978).
- Gadagkar, R. *Proc. Ind. Acad. Sci.* **94**, 587–621 (1985).
- Page, R. E. & Metcalf, R. A. *Am. Nat.* **119**, 263–281 (1982).
- Ross, K. G. *Nature* **323**, 798–800 (1986).
- Rissing, S. W. & Pollack, G. B. in *Interindividual Behavioral Variability in Social Insects* (ed. Jeanne, R. L.) 179–222 (Westview, Boulder, 1988).
- Winston, M. L. *The Biology of the Honey Bee* (Harvard University Press, Cambridge, 1987).
- Laidlaw, H. H. & Page, R. E. *Genetics* **108**, 985–997 (1984).
- Peng, Y., Fang, Y., Xu, S. & Ge, L. *J. Invert. Pathol.* **49**, 54–60 (1987).
- Nixon, H. L. & Ribbands, C. R. *Proc. Roy. Soc.* **140**, 43–50 (1952).
- Free, J. B. *Br. J. Anim. Behav.* **5**, 41–47 (1957).
- Seeley, T. D. *Behav. Ecol. Sociobiol.* **11**, 287–293 (1982).
- Page, R. E., Kimsey, R. B. & Laidlaw, H. H. *Experientia* **40**, 182–184 (1984).
- Milum, V. G. *Proc. 10th Int. Congr. Entomol.* **4**, 1085–1088 (1956).
- Korst, P. J. A. M. & Velthuis, H. H. W. *Insectes Soc.* **29**, 209–221 (1982).
- Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. *Discrete Multivariate Analysis: Theory and Practice* (MIT Press, Cambridge, 1975).
- Winston, M. L. & Katz, S. J. *Behav. Ecol. Sociobiol.* **10**, 125–129 (1982).
- Collins, A. M., Rinderer, T. E., Harbo, J. R. & Bolten, A. B. *Science* **218**, 72–74 (1982).
- Rothenbuhler, W. C. *Amer. Zool.* **4**, 111–123 (1964).
- Hellmich, R. L., Kulincevic, J. M. & Rothenbuhler, W. C. *J. Heredity* **76**, 155–158 (1985).
- Calderone, N. W. & Page, R. E. *Behav. Ecol. Sociobiol.* (in the press).
- Crozier, R. H. & Page, R. E. *Behav. Ecol. Sociobiol.* **18**, 105–115 (1985).
- Sherman, P. W., Seeley, T. D. & Reeve, H. K. *Am. Nat.* (in the press).
- Crozier, R. H. & Consul, P. C. *Theor. pop. Biol.* **10**, 1–9 (1976).
- Kolmes, S. A. *J. Kans. Entomol. Soc.* **58**, 413–421 (1985).
- Winston, M. L. & Fergusson, L. A. *Can. J. Zool.* **63**, 777–780 (1985).
- Moritz, R. F. A. & Southwick, E. E. *Behav. Ecol. Sociobiol.* **21**, 53–57 (1987).

- Robinson, G. E. & Page, R. E. *Nature* **333**, 356–358 (1988).
- Cohen, A. C. Jr *Biometrics* **16**, 203–211 (1960).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1981).
- Tucker, K. W. in *Bee Genetics and Breeding* (ed. T. E. Rinderer) 57–90 (Academic, Orlando, 1986).

Clutch size adjustment by a swallowtail butterfly

Diana Pilson & Mark D. Rausher

Department of Zoology, Duke University, Durham, North Carolina 27706, USA

Because life-history characters such as breeding schedule, reproductive investment, and age-specific survivorship exhibit great variation in nature and are closely linked to individual fitness, the development of a theory of life-history evolution is a major focus of evolutionary biologists^{1–3}. One life-history trait that has received much attention is clutch size, the number of eggs laid by a female during a round of reproduction. For parasitic organisms such as herbivorous insects and parasitoids, two models have been developed which describe the evolution of behaviour influencing clutch size^{4,5}. These models both make two predictions: (1) that individuals should adjust clutch size so that larger clutches are laid on hosts that can support the growth of more offspring, and (2) that as time between oviposition bouts increases, as would occur when hosts become rarer, larger clutches should be laid on hosts of a given quality. We present here the first empirical test of these predictions and show that the pipevine swallowtail butterfly, *Battus philenor*, adjusts its clutch size in response to variation in both host quality and time since last oviposition.

We performed an experiment in the field to determine whether females modify clutch size in response to variation in host-plant characters and lay larger clutches on plants of higher quality. If characteristics of the host plant, *Aristolochia reticulata* (Aristolochiaceae), do not influence clutch size, there should be no correlation between clutch size and any plant characters. But clutch size was significantly correlated with most plant characters we measured, suggesting that females modify clutch size depending on the characteristics of the plant (Table 1). Because young plants have fewer, smaller leaves and longer, wider buds than old plants and *A. reticulata* leaves become sclerophyllous and less edible as they age⁶, the observed correlations suggest that females lay larger clutches on younger plants with more edible foliage.

This conclusion is supported by examination of the size at which test larvae disperse from host plants. We removed original clutches from plants and then seven days later (normal egg

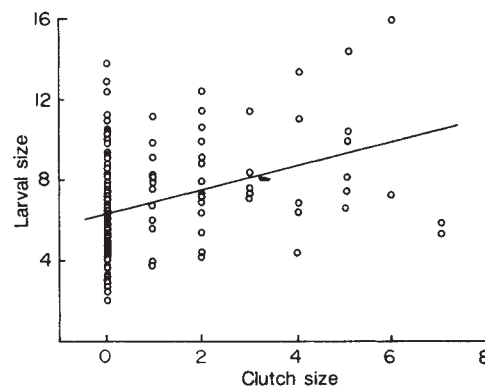


Fig. 1 Correlation between egg cluster size and mean length at dispersal of test larvae on each plant ($r = 0.33$, $P < 0.0001$, $n = 163$). Laboratory data show that both plant quality and time since last oviposition affect clutch size. In the field experiment time since last oviposition was not controlled and this is probably one reason for scatter in the data presented here. For methods see Table 1.

Table 1 Correlations between plant and butterfly characters.

Plant character	Clutch size	Mean length of test larvae
Bud length	$r = 0.15$ $P < 0.06$ $n = 162$	$r = 0.36$ $P < 0.0001$ $n = 162$
Bud width	$r = 0.29$ $P < 0.009$ $n = 79$	$r = 0.53$ $P < 0.0001$ $n = 79$
Number of leaves	$r = -0.27$ $P < 0.0005$ $n = 162$	$r = -0.29$ $P < 0.0002$ $n = 162$
Leaf area	$r = -0.28$ $P < 0.0002$ $n = 163$	$r = -0.15$ $P < 0.06$ $n = 163$

Correlations between plant characters, clutch size, and mean length of two test larvae at dispersal. (r = correlation coefficient; P = probability level; n = sample size, number of clutches.) During March and early April 1980 at the John Henry Kirby State Forest in Tyler County, Texas, USA, we followed females searching for their host plant, *Aristolochia reticulata* (Aristolochiaceae). This species has determinant growth and only produces new foliage during a short period in the early spring. As the foliage matures it becomes sclerophyllous and a poor substrate for larval growth. A single plant rarely contains enough young foliage to support the complete growth and development of a larva. Consequently, when larvae have consumed all edible foliage on a plant, they disperse to new ones. The probability that a larva will survive to discover a new host plant is proportional to its size at dispersal⁷. For each host plant on which a female alighted we recorded the number of eggs laid, removed any eggs laid, and measured terminal bud length, bud width, number of leaves, and total leaf area⁸. Seven days later, the normal duration of egg development in the field, we placed two randomly chosen newly hatched test larvae on each host plant. We censused the plants daily and recorded the length of each remaining larva and the data and cause of disappearance of each missing larva⁸. Larval length and mass are highly correlated⁷. We used only plants on which both larvae disappeared by dispersal because for these plants the mean size of the two test larvae should reflect the amount of edible foliage on the plant.

development time) placed two randomly chosen newly hatched larvae on the plants. Mean size at dispersal of the two test larvae increased with increasing cluster size (Fig. 1), indicating that females tend to place larger clusters on plants with more edible foliage, supporting prediction (1) above. As expected, larval size at dispersal, like egg cluster size, is negatively correlated with plant size, and positively correlated with bud size (Table 1).

We performed a second experiment in the laboratory to test both predictions under more controlled conditions than was possible in the field. Our rationale was to control both the quality of a host plant on which a female was allowed to lay eggs and also the readiness of a female to do so, which is expected to increase with time since the last oviposition. All females tested progressed through three distinct levels of discrimination (see legend to Table 2). Females rejected all plants, regardless of food quality, immediately after laying eggs. After this initial rejection period, females accepted high-quality plants but rejected those of low quality. Finally, after this discrimination phase, females accepted both types of plant. Females thus consistently discriminate between high- and low-quality *A. reticulata* plants in a way very similar to the way butterflies discriminate among different host species⁹.

Females at the same discrimination level presumably represent those with a similar readiness to lay eggs. By comparing the mean clutch size of such females on hosts of different quality, we tested whether host quality affects clutch size independently of readiness to lay eggs. Such an effect exists for *B. philenor*: for females that have just accepted both the high- and low-quality plants, mean clutch size is significantly larger on high-quality plants than on low-quality plants (Table 2). This parallels

our field results, indicating that females tend to lay larger clutches on plants with young, edible foliage, and supports prediction (1).

By allowing females with different discrimination levels to lay eggs on plants of the same quality, the effect on clutch size of a female's readiness to lay eggs can be assessed independently of host quality. Females that have just accepted both high- and low-quality plants lay larger clutches on high-quality plants than females that have only accepted high-quality plants (Table 2). This result implies that the longer a female goes without laying eggs on a host plant, the larger her clutch size will be when she does so, supporting prediction (2).

Insects discriminate between different host plant species when deciding to lay eggs¹⁰⁻¹⁸ and a female's discrimination level may progress through well defined stages⁹. Our results indicate that this kind of progression also characterizes discrimination between individual plants within a host species. In *B. philenor* host-plant selection involves first locating an individual of an appropriate host and then, depending on the quality of the plant and the discrimination level of the female, determining a clutch size ranging from zero to some maximum. Our results suggest that there is some physiological variable that increases with time since last oviposition but decreases with each egg laid. One commonly suggested candidate for a physiological variable with these properties is egg load¹⁹, which could be monitored by stretch receptors in the abdomen.

Table 2 Average egg cluster size in each treatment.

Quality of plant	Discrimination level	
	Accept high quality plant only (1)	Accept both plants (2)
High	2.73 (0.43)	7.27 (1.14) ($P < 0.006$)
Low		3.20 (0.46)

P -values indicate significance level of difference between treatments (Mann-Whitney U -test; $n = 15$ for all treatments). Numbers in parentheses are standard errors. In the summer of 1985 in our laboratory in Durham, North Carolina, USA, mated females were allowed to lay eggs on a high-quality *A. reticulata* plant until they refused to accept that plant for oviposition. This procedure reduced all females to the same low level of readiness to lay eggs. Before testing, each female was randomly assigned to one of three treatments. In treatment 'high-1', females were permitted to lay eggs on a high-quality plant as soon as they accepted the high-quality plant but before they accepted the low-quality plant. In treatment 'high-2', females were permitted to lay eggs on the high-quality plant as soon as they accepted both the high- and low-quality plants. Finally, females in treatment 'low-2' were allowed to lay eggs on the low-quality plant as soon as they accepted both the high- and low-quality plants. Plant appearance changes dramatically with age. High-quality plants had six young tender light green leaves and two large terminal buds. Low-quality plants had six old sclerophyllous darker green leaves and no terminal buds. Forty-seven females were used in this experiment. All 47 accepted the high-quality plant before accepting the low-quality plant. Of these 47, 15 had been assigned to treatment high-1 and were allowed to lay eggs immediately. Of the remaining 32, 30 later accepted the low-quality plant and then were allowed to lay eggs on either the low-quality plant (treatment low-2) or the high-quality plant (treatment high-2). The remaining two females never accepted the low quality plant and were removed from the experiment. Following the methods of Singer⁹, after a female's readiness had been zeroed, females were placed alternately on a high- and low-quality *A. reticulata* plant every 4-5 minutes. A female was considered to have accepted a plant on which she was placed if she curled her abdomen to the foliage in an attempt to lay eggs, since this behaviour is almost invariably followed by oviposition. If continued testing of the female was desired, she was removed from the plant before any eggs were laid. A female was considered to have rejected a plant if she either assumed a basking position or flew off the plant.

Most theoretical and empirical investigations of the evolution of clutch size have considered organisms with parental care that lay one clutch per season²⁰⁻²². By contrast, many short-lived organisms such as insects lay several clutches within a season and then die. Although in these organisms the variation in clutch size is large, both between species and among individuals^{10,23-26}, theory developed for seasonal breeders is of little help in

accounting for this variation. Two recent models, however, specifically address this type of variation in herbivorous and parasitic insects^{4,5}. Our results substantiate two of the predictions of these models, suggesting that their underlying assumptions are valid for at least one species of herbivore.

We thank W. Cook, P. Feeny, L. Fleishman, D. Papaj and A. Weakley. This work was supported by an NSF grant to M.D.R.

Received 14 March 1988; accepted 20 April 1988.

1. Stearns, S. C. *Q. Rev. Biol.* **51**, 3-47 (1976).
2. Stearns, S. C. *A. Rev. Ecol. Syst.* **8**, 145-171 (1977).
3. Charlesworth, B. *Evolution in Age-Structured Populations* (Cambridge University Press, Cambridge, 1980).
4. Parker, G. A. & Courtney, S. P. *Theor. Pop. Biol.* **26**, 27-48 (1984).
5. Skinner, S. W. *Behav. Ecol. Sociobiol.* **17**, 231-238 (1984).
6. Rausher, M. D. *Ecol. Monogr.* **51**, 1-20 (1981).
7. Rausher, M. D. *Anim. Behav.* **27**, 1034-1040 (1979).
8. Rausher, M. D. & Papaj, D. *Ecology* **64**, 1402-1410 (1983).
9. Singer, M. C. *Oecologia* **52**, 224-229 (1982).
10. Papaj, D. & Rausher, M. D. in *Herbivorous Insects: Host Seeking Behavior and Mechanisms* (ed. Ahmad, S.) 77-124 (Academic, New York, 1983).
11. Waage, J. K. & Ng, S. M. *J. Anim. Ecol.* **53**, 401-415 (1984).

12. Mitchell, J. *Ecology* **56**, 696-702 (1975).
13. Ives, P. M. *Aust. J. Ecol.* **3**, 261-276 (1978).
14. Wolfson, J. L. *Eng. exp. Appl.* **27**, 223-232 (1980).
15. Rausher, M. D., Mackay, D. A. & Singer, M. D. *Anim. Behav.* **29**, 1220-1228 (1981).
16. Zucker, W. V. *Ecology* **63**, 972-981 (1982).
17. Bultman, T. L. & Faeth, S. H. *Oecologia* **69**, 117-120 (1986).
18. Singer, M. C. *Evolution* **37**, 389-403 (1983).
19. Odendaal, F. J. *J. Insect Behav.* (in the press).
20. Lack, D. *Ibis* **89**, 302-352 (1947).
21. Lack, D. *The Natural Regulation of Animal Numbers* (Oxford University Press, 1954).
22. Stonehouse, B. & Perrins, D., eds *Evolutionary Biology* (Macmillan, London, 1977).
23. Chew, F. S. & Robbins, R. K. *Symp. R. Ent. Soc. Lond.* **11**, 65-79 (1983).
24. Courtney, S. P. *Am. Nat.* **123**, 276-281 (1984).
25. Hayes, J. L. *Oecologia* **66**, 495-498 (1985).
26. Stamp, N. E. *Am. Nat.* **115**, 367-380 (1980).

Early vision and texture perception

James R. Bergen* & Edward H. Adelson†

*SRI David Sarnoff Research Center, Princeton, New Jersey 08540, USA

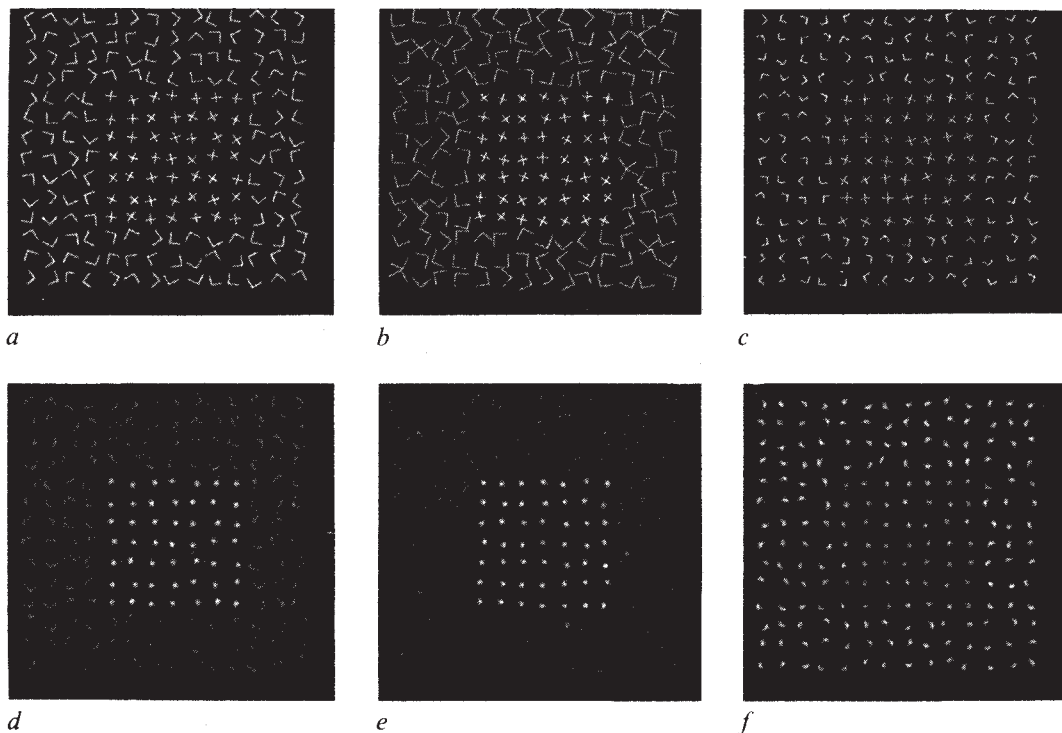
†Media Lab and Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Texture perception has frequently been studied using textures constructed by repeated placement of micropatterns or texture elements. Theories have been developed to explain the discriminability of such textures in terms of specific features within the micropatterns themselves. For example, Beck^{1,2} observed that a region filled with vertical Ts is readily distinguished from one filled with tilted Ts but not from one filled with vertical Ls. He attributed this to the different distribution of oriented line segments present in the former case but not in the latter. However, Bergen and Julesz³ found that a region of randomly oriented Xs segregated from one filled with randomly oriented Ls, in spite of the identical

distribution of oriented line segments in the two cases. They suggested that this discrimination might be based on the density of such features as terminators, corners, and intersections within the patterns. We note here that simpler, lower-level mechanisms tuned for size may be sufficient to explain this discrimination. We tested this by varying the relative sizes of the Xs and the Ls; when they produce equal responses in size-tuned mechanisms they are hard to discriminate, and when they produce different size-tuned responses they are easy to discriminate.

Figure 1a shows an example of a texture composed of Xs within a texture composed of Ls, similar to the texture used by Bergen and Julesz. The X and L micropatterns are each made of two perpendicular bars, and the bars making the Xs have the same length and thickness as the bars making the Ls. The textures are easily distinguished, and Bergen and Julesz suggested that discrimination could be accomplished by mechanisms that measured the differing densities of micropattern features; for example the Xs have four terminators each, while the Ls only have two; the Ls have a corner while the Xs do not; and so on. This type of description was motivated by analysis of many textures constructed of small micropatterns made up of line segments, dots or other discrete components.

Fig. 1 Top row, Textures consisting of Xs within a texture composed of Ls. The micropatterns are placed at random orientations on a randomly perturbed lattice. *a*, The bars of the Xs have the same length as the bars of the Ls. *b*, The bars of the Ls have been lengthened by 25%, and the intensity adjusted for the same mean luminance. Discriminability is enhanced. *c*, The bars of the Ls have been shortened by 25%, and the intensity adjusted for the same mean luminance. Discriminability is impaired. Bottom row, The responses of a size-tuned mechanism. *d*, response to image *a*; *e*, response to image *b*; *f*, response to image *c*.



One major difficulty with this approach is that it is based on a verbal description of image features rather than on the raw intensity values in the image itself. This makes it difficult to test under more general conditions. In order to apply this analysis to a more general class of images, it would first be necessary to construct operators that extract the feature descriptions being invoked—a task that has yet to be accomplished. Before embarking on such a difficult approach it is worth asking whether simpler extracted properties, such as those derivable from linear filters, will suffice (see refs 4–7).

When inspecting this texture one may observe that the Xs look smaller than the Ls, and that they break up the background differently. This suggests that very simple size-tuned mechanisms, such as cells with centre-surround receptive fields, could play an important role in the discrimination. We changed the relative sizes of the Xs and the Ls to see whether we could increase and decrease the discriminability of the patterns.

Figure 1b shows the result of lengthening the bars of the Ls by 25%. The bar intensities have been compensated so that the overall density of the micropatterns (that is, the equivalent amount of ink in each) is unchanged. The discrimination becomes easier. Thus, although the micropatterns still have the same number of terminators, corners, and so on, the manipulation of size has a significant impact on the discriminability of the texture.

Figure 1c shows the result of making the bars of the Ls 25% shorter than in the original textures, again with compensation in the intensity of the bars. Now the discrimination is more difficult.

Figure 1d–f shows the response of perhaps the simplest size-tuned mechanism we can construct: a linear centre-surround receptive field followed by full-wave rectification. Figure 1d shows the response to the stimulus of Fig. 1a; the mechanism responds more strongly to the patch in the centre. Figure 1e shows the response to Fig. 1b; now the differences are even more apparent. Figure 1f shows the response to Fig. 1c; in this case the size-tuned mechanism gives responses of similar strength to the two textures.

For this particular set of textures, then, the discriminability can be predicted fairly well from the activities of size-tuned units, without reference to more feature-like properties of the micropatterns. We suggest that the visual system uses a two-stage cascade of local energy measures (similar to the cascade of orientation measures discussed by Knutson and Granlund⁵.) In the first stage, linear filters are followed by a rectifying nonlinearity (as in fig. 1d–f); spatial averaging provides primary energy measures. These responses are then treated as image arrays for input to a further layer of linear filters, which compute secondary energy measures that indicate the locations of texture boundaries.

Models for texture perception that are based on concepts such as 'terminators' and 'corners' have been important in motivating research in early vision, but the models have proven difficult to formalize in such a way that they can be applied to wide classes of textures. Although we do not present a full model of texture perception here, the above demonstration indicates that simple filtering processes operating directly on the image intensities can sometimes have surprisingly good explanatory power. The accompanying paper by Voorhees and Poggio⁸, based on a computational investigation into texture analysis, offers an example of a more fully elaborated theory and further demonstrates the potential power of simple processes in early vision.

5. Knutson, H. & Granlund, G. H. (1983) *IEEE Workshop of Computer Architecture for Pattern Analysis and Image Data Base Management* 12–14, Pasadena (IEEE Computer Society Press, Washington DC, 1983).
6. Voorhees, H. & Poggio, T. *Proceedings, First International Conference on Computer Vision*, 250–258, London (IEEE Computer Society Press, Washington DC, 1987).
7. Bergen, J. R. & Adelson, E. H. *J. opt. Soc. Am.* A3, 99 (1986).
8. Voorhees, H. & Poggio, T. *Nature* 333, 364–367 (1988).

Computing texture boundaries from images

Harry Voorhees* & Tomaso Poggio

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Recent computational and psychological theories of human texture vision^{1–3} assert that texture discrimination is based on first-order differences in geometric and luminance attributes of texture elements, called 'textons'⁴. Significant differences in the density, orientation, size, or contrast of line segments or other small features in an image have been shown to cause immediate perception of texture boundaries. However, the psychological theories, which are based on the perception of synthetic images composed of lines and symbols, neglect two important issues. First, how can textons be computed from grey-level images of natural scenes? And second, how, exactly, can texture boundaries be found? Our analysis of these two issues has led to an algorithm that is fully implemented and which successfully detects boundaries in natural images⁵. We propose that blobs computed by a centre-surround operator are useful as texture elements, and that a simple non-parametric statistic can be used to compare local distributions of blob attributes to locate texture boundaries. Although designed for natural images, our computation agrees with some psychophysical findings, in particular, those of Adelson and Bergen (described in the preceding article⁶), which cast doubt on the hypothesis that line segment crossings or termination points are textons.

How can texture elements be computed from images of natural scenes? From the grey-level image our algorithm extracts 'blobs'—small compact and elongated linear regions (also known as 'bars') which are darker or lighter than their surround. To detect relatively dark blobs, the algorithm works as follows. First, the image is convolved with a centre-surround filter at a fine scale (we typically use a laplacian of gaussian of standard deviation 1.5 pixels). By using the logarithm of intensity, the resulting blob regions are fairly insensitive to shadows or other large-scale changes in illumination. Positive values of the filtered image indicate regions which are relatively darker than their surround. Hence, blobs can be regarded as the duals of intensity edges, which are often computed as the zero crossings of the filtered image. To remove some spurious connections between blob regions, the filtered image is thresholded about a small positive value. The threshold is proportional to an estimate of noise in the image, which is automatically computed from a histogram of the intensity gradient magnitude⁵.

This computation yields a map of dark blob pixels. Using morphological operations, these areas are segmented on the basis of local shape into small compact and thin elongated components. For each of these blobs, five attributes are computed—contrast, orientation, width, length, area and area density. Finally, spurious or noisy blobs are removed by requiring that the intensity gradient at most points along a blob's perimeter are above a noise threshold and pointing outwards.

We have found that in natural images, the computed blobs and bars capture physically meaningful information about the surface—small impressions and markings—and that attributes

Received 20 January; accepted 26 February 1988.

1. Beck, J. *Percept. Psychophys.* 1, 300–302 (1966).
2. Beck, J. *Percept. Psychophys.* 2, 491–495 (1967).
3. Bergen, J. R. & Julesz, B. *Nature* 303, 696–698 (1983).
4. Caelli, T. *Spatial Vision* 1, 19–30 (1985).

* Present address: Thinking Machines Corporation, 245 First Street, Cambridge, Massachusetts 02142, USA.

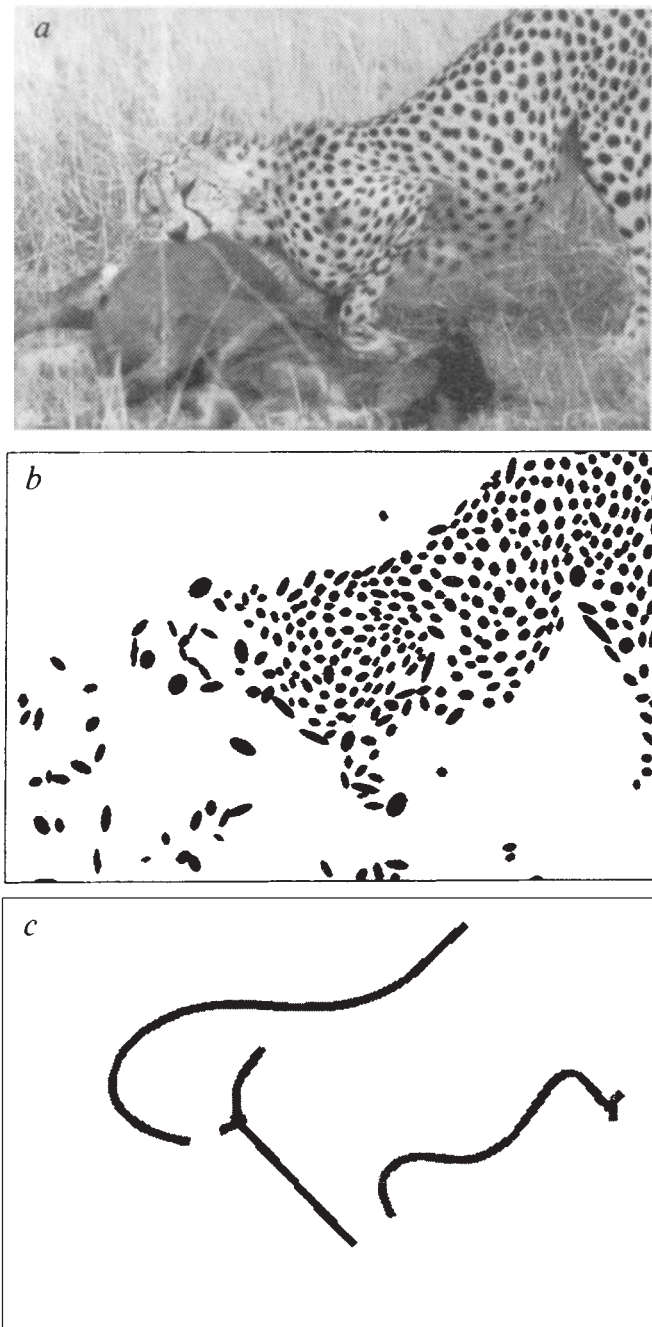


Fig. 1 *a*, Image of a leopard (from *Time*; 23 February 1987, page 40). *b*, The computed textons (blobs and bars). *c*, Texture boundaries due to density as found by our algorithm.

of these blobs characterize the texture. The algorithm, therefore, provides a method of computing discrete 'textons' from natural images in the spirit of Marr's raw primal sketch¹. Figures 1 and 2 show the result of applying our algorithm to natural images of textured surfaces.

Once blobs are computed, how can their attributes be compared to locate texture boundaries? Specifically, what statistic is used to compare two distributions of attributes from nearby neighbourhoods? We define a non-parametric statistic which reliably predicts under which circumstances a texture boundary will be perceived between two regions.

Mimicking the inability of humans to immediately discriminate small changes in attribute values, we use histograms rather than the exact sample distributions themselves. The histogram bucket size equals the minimum difference which is immediately discriminable (for example, 15° for orientation). From the stand-

point of detecting physical boundaries, implementing this 'weakness' makes sense: if we did not ignore small differences in attribute values, a graded texture gradient, perhaps formed by the projection of a curved surface, would yield undesirably significant texture boundaries across its face. Discretizing the distribution (with overlapping buckets) avoids this problem.

To compare two sample attribute distributions, we compute the maximum difference between corresponding histogram buckets. This maximum frequency difference (MFD) statistic is better suited than most standard statistics for this purpose, because it is less sensitive to outliers, which are common in natural images. The statistic also agrees with several human discrimination capabilities. For example, a region of randomly oriented bars is readily distinguished from a region of equally oriented bars, and the two distributions yield a high MFD. Conversely, we were surprised to discover that a region of bars randomly oriented between 0 and 90° is hard to discriminate from a region of bars randomly oriented between 90 and 180°. Any standard statistical test would regard these distributions as very significantly different. The MFD, however, is small, again in accord with perception⁵.

A ubiquitous problem in texture discrimination is the choice of neighbourhood size. It has been said that large neighbourhoods are needed to robustly represent texton distributions in the presence of noise and outliers, but that small neighbourhoods are needed to localize or even to avoid missing certain texture boundaries. Our statistic seems to offer a way out of this dilemma. Because it is quite insensitive to outliers, small distributions (of about ten samples) often suffice to compute reliable difference measures. Such neighbourhoods are small enough to detect boundaries between regions which are as narrow as two or three textons across. In our implementation, the neighbourhood diameter is between 5 and 10 times the average blob width. These figures are based on some psychophysical results of Julesz⁷, and are designed so that any neighbourhood of textons dense enough to be perceived as a texture is of sufficient size.

When applied to the synthetic images tested by Bergen and Adelson⁶ in the previous article, our computation provides an alternative to the conjecture³ that crossings are textons (see Fig. 3). In the left image, originally offered as evidence of crossings as textons, the crosses give smaller blobs of higher contrast than do the *L*s (Fig. 3*b*). This demonstrates that blob attributes offer an alternative to crossings as an explanation of human texture discrimination.

In the right image, where the size of the *L*s and the intensity of crosses are decreased, the computed blobs from each region are similar. Bergen and Adelson⁶ reported in this case that discrimination is harder. It therefore appears that simple blob attributes offer not only an alternative, but a more accurate explanation of human discrimination capabilities. At the very least, the psychophysical results demonstrate that such spatial filtering cannot be neglected in a satisfactory explanation of human texture perception. Recent psychophysical experiments⁸ also support this view.

We go on to illustrate how our algorithm computes texture boundaries in this example. Figure 3*c* shows histograms of blob contrast values from neighbourhoods in different regions in each of the two images, together with the computed MFD statistic. Figure 3*d* shows the MFD computed at a grid of points across each image array. Significantly high and long ridges in each array identify blob contrast boundaries, as shown in the left image of Fig. 3*e*. In the right image, no significant boundaries are detected. Similar results are obtained on this example using blob width as an attribute instead of blob contrast.

The boundaries computed by our algorithm are consistent with pre-attentive texture perception, without the need for crossings as textons. From the standpoint of analysing natural images, this result is not surprising. Seldom would a natural scene give texture elements shaped like characters, each having a distinct number of crossings or terminators. Although we have not

Fig. 2 *a*, Natural image of clothing. *b*, The computed textons (blobs and bars). *c*, Texture boundaries found by our algorithm. Boundaries due to differences blob width are shown in white with boundaries due to differences in blob contrast shown in grey. The algorithm has been applied to several other examples of natural texture⁵.

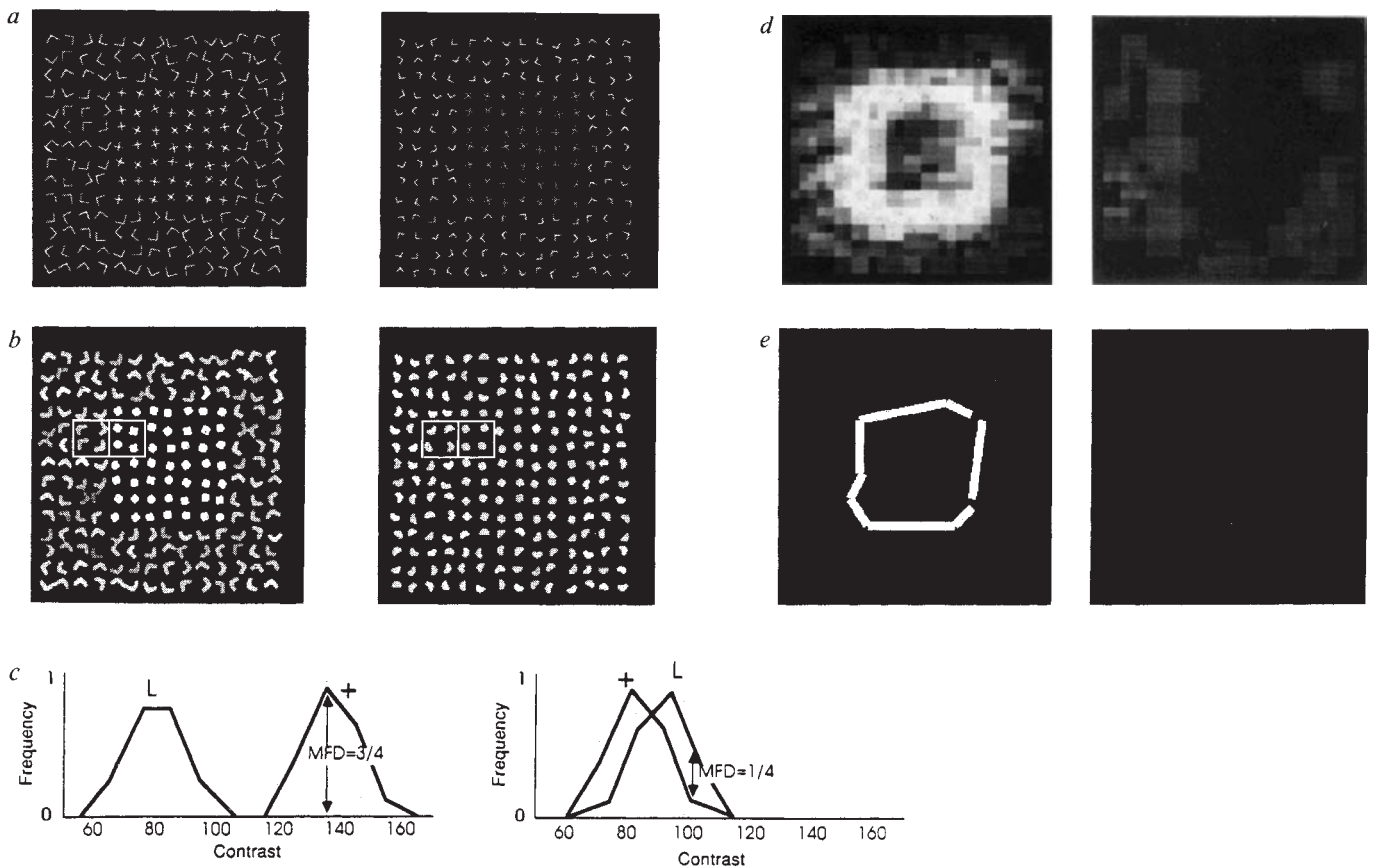
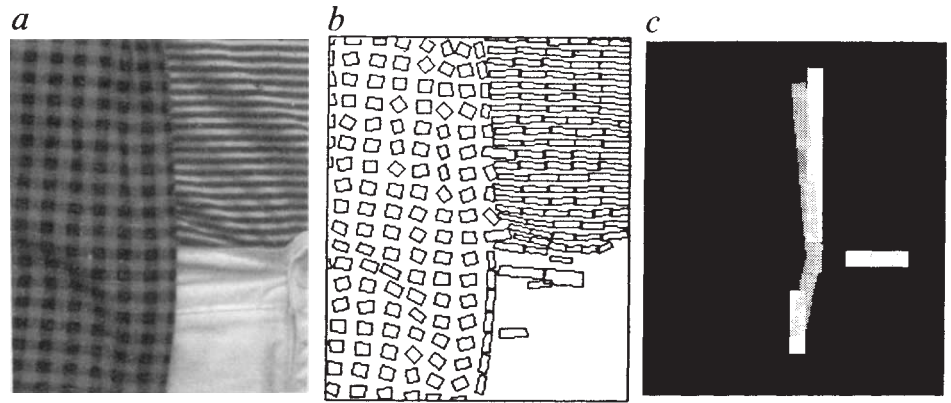


Fig. 3 Demonstration that crossings are not textons. *a*, (Left) Demonstration offered as evidence that crossings are textons. (Right) Bergen and Adelson's example⁶ where discrimination is decreased. *b*, Blobs obtained by $\nabla^2 G$ filtering, shown here with intensity proportional to blob contrast. At left, the crosses give smaller blobs of higher contrast than do the *L*s. The gaussian filter, *G*, has standard deviation 2.5 pixels, and the image size is 256 pixels square. *c*, Superimposed histograms of blob contrast from adjacent neighbourhoods shown in *b*, with maximum frequency difference (MFD) statistic. In our implementation, the histograms are smoothed slightly to reduce the effect of discretization. *d*, MFD of contrast, computed from adjacent neighbourhoods over each entire array. *e*, Significantly high and long ridges of *d* identify contrast boundaries. Contrast accounts for the discrimination of the left image, without the explicit computation of crossings. No boundaries are detected in the right image, which humans also find harder to discriminate according to Bergen and Adelson (see preceding article⁶).

proven that blobs are textons any more than Julesz and Bergen³ proved that crossings are textons, we believe that our explanation is preferable, because the computation works for natural images as well as for synthetic ones. Also, in natural images, blob attributes provide information about the physical cause of a texture boundary. Changes only in blob orientation, for example, suggest the presence of a surface orientation discontinuity, while large changes in size or density can only be due to occlusion or to a change in material⁹.

Our experience with natural images suggests that most texture boundaries are probably due to differences in texton density. At a particular scale of analysis, two surfaces adjacent in the

image are unlikely to give textons at the same scale of analysis (as in Fig. 1). It is possible that texture boundaries due only to changes in texton orientation or size (as in Fig. 2) are less common in images of natural scenes than synthetic, psychophysical images would lead us to believe.

We suspect that terminators may not be textons either. We have already shown how the discrimination of Fig. 3, which contains terminator density ratio of 4:2, can be explained by blob attributes. We also found, using images of squarish randomly oriented 6s and Es (similar to those of Fig. 6 of Julesz⁴), that even a terminator density ratio of 3:1 is not sufficient for discrimination. Breaks in lines (pairs of terminators) seem more

salient than ends of lines, and may be more relevant to natural textures (see ref. 5, Fig. 6.5). Breaks in dark lines may be computed as small-scale light blobs.

Our computation leaves many questions unanswered. We have not dealt with multiple scales of blob detection or the integration of various attribute boundaries. Nevertheless, we believe that we have extended perceptually based theories of texture vision to natural images by implementing an algorithm which computes textons, and which compares textons to locate texture boundaries, in accordance with human texture capabilities. We have successfully applied the algorithm to natural images, and have shown how it accounts for previous psychophysical results as well. Our algorithm demonstrates the feasibility of a first-order, symbolic approach to texture discrimination in the spirit of Marr's primal sketch¹ and Julesz's texton theory⁴. It is possible that our statistical test applied to the outputs of a sufficient number of arrays of linear filters would also provide acceptable texture segmentation. Such a scheme, similar to Bergen and Adelson's suggestion⁶, remains to be elaborated, implemented and tested. It is possible that the main steps of the algorithm could be implemented in a relatively natural way by known neural structures in the retina and the visual cortex.

The authors thank Fero Simoncelli and Mike Sokolov for their help in producing the images.

Received 21 January; accepted 7 April 1988.

1. Marr, D. *Cold Spring Harb. Symp. quant. Biol.* Vol. XL, 647-662 (1976).
2. Beck, J., Prazdny, K. & Rosenfeld, A. in *Human and Machine Vision* (eds Beck, J., Hope, B. & Rosenfeld, A.) 1-38 (Academic, New York, 1983).
3. Julesz, B. & Bergen, J. R. *The Bell System Technical Journal* 62, 1619-1645 (1983).
4. Julesz, B. *Nature* 290, 91-97 (1981).
5. Voorhees, H. *M.I.T. Artificial Intelligence Lab, Technical Report number 968*, 1987.
6. Bergen, J. P. & Adelson, E. H. *Nature* 333, 363-364 (1988).
7. Julesz, B. *Biol. Cybern.* 54, 245-251 (1986).
8. Beck, J., Sutter, A. & Ivry, R. *Computer Vision Graphics Image Proc.* 37(2), 299-325 (1987).
9. Riley, M. D. *M.I.T. Artificial Intelligence Lab, Technical Report 649* (1981).

A cellular analogue of visual cortical plasticity

Yves Frégnac, Daniel Shulz, Simon Thorpe*
& Elie Bienenstock

Laboratoire de Neurobiologie et Neuropharmacologie du Développement, Bâtiment 440, Université Paris XI, F-91405, Orsay Cedex France

Neuronal activity plays an important role in the development of the visual pathway. The modulation of synaptic transmission by temporal correlation between pre- and postsynaptic activity is one mechanism which could underlie visual cortical plasticity¹⁻⁴. We report here that functional changes in single neurons of area 17, analogous to those known to take place during epigenesis of visual cortex^{5,6}, can be induced experimentally during the time of recording. This was done by a differential pairing procedure, during which iontophoresis was used to artificially increase the visual response for a given stimulus, and to decrease (or block) the response for a second stimulus which differed in ocularity or orientation. Long-term modifications in ocular dominance and orientation selectivity were produced in 33% and 43% of recorded cells respectively. Neuronal selectivity was nearly always displaced towards the stimulus paired with the reinforced visual response. The largest changes were obtained at the peak of the critical period in normally reared and visually deprived kittens, but changes were also observed in adults. Our findings support the role of temporal correlation between pre- and postsynaptic activity in the induction of long-lasting modifications of synaptic transmission during development, and in associative learning.

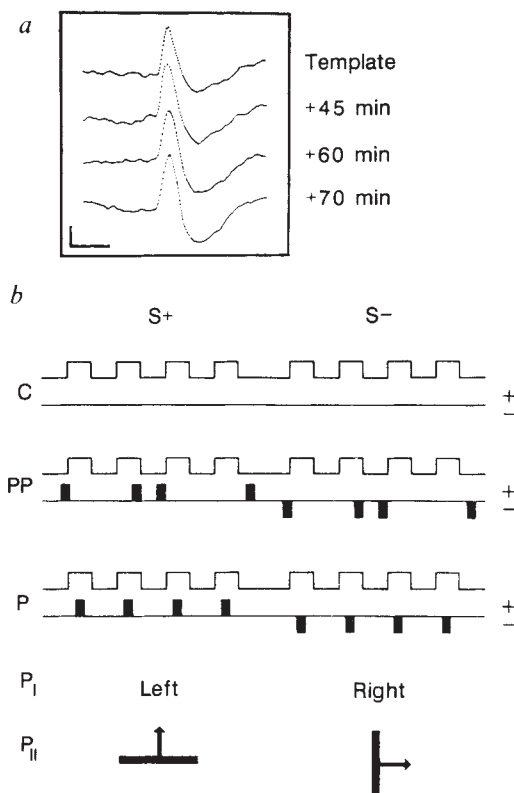


Fig. 1 Imposed temporal correlation between visual afferent activity and postsynaptic firing induced iontophoretically. *a*, For each cell, the shape of the action potential (calibration bars, 1 mV and 1 ms) was continuously monitored on a digital scope to ensure that the same neuron was recorded throughout the experiment. *b*, Each indentation indicates the temporal occurrence of a single visual stimulation, and each filled rectangle that of an iontophoretic pulse of a given polarity (shown at right). During control (C, upper row), two stimuli (S⁺ and S⁻) were presented by blocks of four trials in succession (upper line), without iontophoretic current (lower line). During pseudo-pairing (PP, middle row), iontophoretic pulses were uncorrelated with visual stimulation. During pairing (P, lower row) iontophoretic pulses were concomitant with the visual response, in such a way as to impose a significant increase (S⁺) or decrease (S⁻) of the visual response. Cases where no modulation of activity could be produced by iontophoresis are noted S⁰ (see right eye stimulation in Fig. 2). The two test stimuli differed in either ocular dominance (P_I), or orientation (P_{II}). The relative preference response between the two test stimuli, given by the normalized ratio of visual responses S⁺/(S⁺+S⁻), was measured during control periods in which the sequence of stimulation shown in row C was repeated 10-50 times. The two series of values taken from this ratio before and after pairing (or pseudo-pairing) were compared using both parametric (unpaired Student's *t*-test; significance level of *P* < 0.005) and non-parametric tests (Kolmogorov-Smirnov [K.S.]; significance level of *P* < 0.05). To assess non-associative effects, pseudo-pairing procedures (PP) were interposed between controls prior to pairing in some experiments. In the case of the orientation protocol (P_{II}), the full orientation tuning curves were analysed before and after pairing, to reveal possible generalization effects for stimuli other than those used during pairing.

Methods. In addition to standard anaesthesia and electrophysiological procedures detailed elsewhere¹⁶, we used a simple method to artificially control the level of postsynaptic activity by varying the retention/ejection current of the 1-3 M potassium acetate or chloride extracellular recording electrode (2-20 MΩ). Positive current (average value of +4 nA) and increase in the concentration of potassium in the extracellular medium resulted (in 79% of the cases) in a significant increase in spontaneous and/or evoked activity. Negative current (average value of -9 nA) reduced the cells' activity through a field effect¹⁷ (in 76% of the cases) which sometimes led to a total blockade of the response for the preferred stimulus (in 16% of visual cells).

* Present address: Institut des Neurosciences, University Paris VI, F-75230, Paris, France.

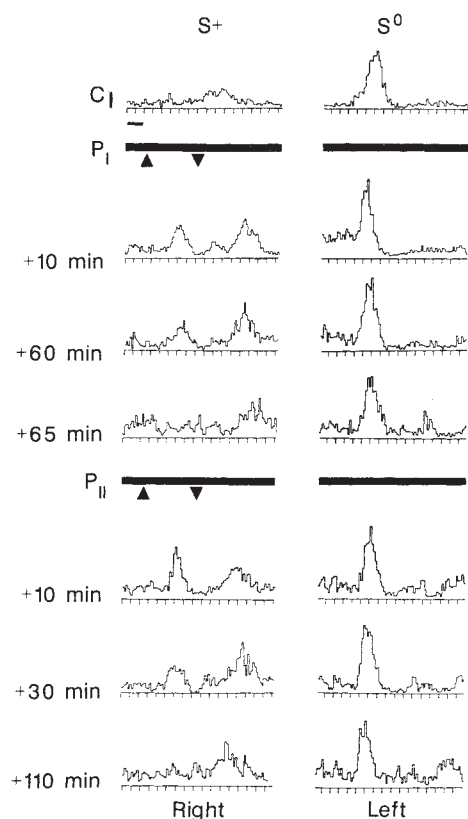


Fig. 2 Ocular dominance change. Recordings were made from a cell in area 17 of a 31-day-old kitten reared normally. PSTHs represent visual response to a bar drifting across the receptive field, using the preferred orientation, through the left eye (left column) and right eye (right column), before and after two pairing periods (thick black lines). Recordings were made for 40 presentations to each eye, interposed in blocks of four as described in Fig. 1. The cell was initially dominated strongly by the right eye (C). During each pairing procedure (not shown), left eye stimulation (S^+) was associated with a +5 nA pulse of current (filled triangles indicate the onset and offset of iontophoresis). No iontophoretic current was applied during stimulation of the right eye (S^0). The first pairing procedure (9 paired stimulations for each eye) resulted in an increase of the visual response to the left eye (+40%) within a restricted zone of the receptive field, and this effect was retained for 60 min (ratio $S^+/(S^++S^0)$; K.S., $P < 0.002$) after the end of the pairing. With time the ocular dominance ratio values returned to control levels, but a second pairing procedure (24 paired stimulation for each eye) imposed a 90% increase in firing within the same restricted zone. This second enhancement, confined exclusively to the paired zone, was retained for a 100 min following pairing (K.S., $P < 0.032$). The second zone of discharge of the left eye receptive field (whose activation was outside the iontophoretic pulse) and the response to stimulation of the right eye were unchanged. Calibration bars: 10 action potentials (a.p.) per s; 1 s and 1.5°.

A refined version of Hebb's neurophysiological postulate¹ of synaptic plasticity is the covariance hypothesis^{3,7,8}. Changes in levels of covariation between pre- and postsynaptic activity determine the sign and amplitude of the modification of synaptic efficiency. This hypothesis predicts both increases and decreases in synaptic transmission: synapses whose activation is repeatedly correlated with firing of the target cell would be reinforced^{1,9} as the result of positive change in covariance, whereas those which become predominantly silent at the time of postsynaptic firing would lose their efficiency¹⁰ as the result of negative change in covariance. Moreover, a decrease in synaptic efficiency is also predicted in the case where there is a repetitive failure of given synapses to trigger the postsynaptic cell (as the result of a negative change in covariance^{3,8}). Such

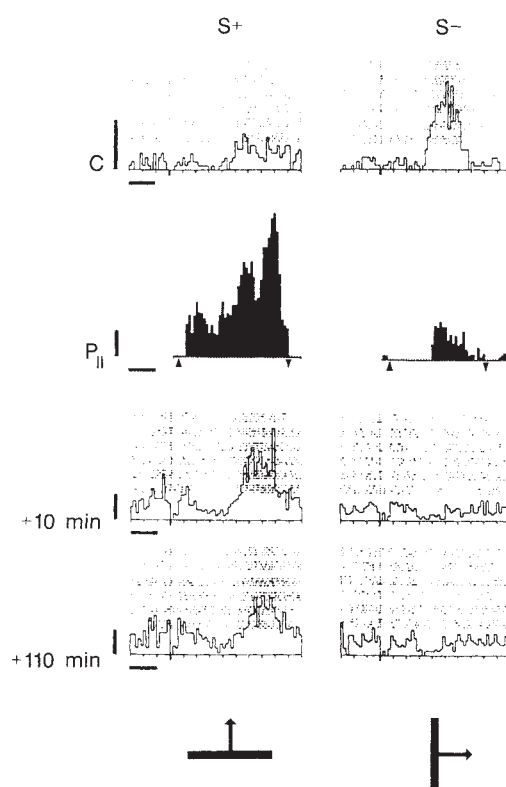


Fig. 3 Orientation preference change. Recordings were made from a cell in area 17 of a 5-week-old kitten reared in the dark for the 3 weeks prior to the experiment. PSTHs represent visual responses to a moving bar (left column, horizontal orientation; right column, vertical orientation). The cell initially showed an orientation bias for a vertical stimulus (upper row, permanent retention current of -3 nA). During pairing (P_{11} , filled histograms), positive current (+3 nA) was applied when the horizontal bar was presented, and interleaved with negative current (-7 nA) when the vertical bar was presented. Iontophoretic pulses produced an enhancement of the visual response for the initially non-preferred stimulus, and a progressive blockade of postsynaptic activity for the initially preferred stimulus which was complete at the end of pairing. The two lower rows show the change in orientation preference observed 10 min (K.S., $P < 0.005$) and 110 min (K.S., $P < 0.005$) after the end of the pairing procedure (60 paired stimulations for each orientation). Calibration bars, 5 a.p. s^{-1} , 1 s and 1.5°. Filled triangles indicate the onset (upwards) and offset (downwards) of iontophoresis.

a rule, which implies competition between active and silent synapses, has been applied to the modelling of the development of neuronal selectivity in visual cortex³.

Functional modifications linked to visual experience have been inferred up to now by comparing distributions of receptive field properties in different populations of neurons in animals raised under different conditions^{5,6,11}. Two intrinsic properties of visual cortical organization, ocular dominance, and orientation selectivity, have been examined in this way and their maintenance or development shown to be experience dependent^{2,5,6,11-14}. By contrast our approach, using electrophysiological methods applied to anaesthetized and paralysed animals, was devised to induce, during the time of recording from a single neuron, functional changes analogous to those involved during epigenesis. Success in this would demonstrate the ability of neurons to change their integrative properties during functioning, and refute the hypothesis that changes in visual cortical function can be explained on the sole basis of disuse and sampling biases in the populations of neurons recorded at different stages of the critical period¹⁵.

We made recordings from 429 cells in area 17 of cats, aged from 4 weeks to adulthood. Controls, performed on 308 of these cells, assessed the invariance of receptive field properties established for the same cell at different times and at different levels of constant iontophoretic current. The other 121 visual cells were submitted to 202 differential pairings and 33 pseudo-pairings (see Fig. 1).

Changes reached levels of statistical significance in the ocular dominance protocol (P_i ; see Fig. 2) in 33% of paired neurons, and they corresponded to a shift in binocularity of 1–2 classes measured on a 5-class scale¹⁶. In 96% of cases this shift was in favour of the eye whose stimulation was paired with the 'high' level of activity. Changes could be produced both in kittens and in adults. Depending on the cell, changes could last from 15 minutes up to at least several hours. Additional pairings could restore or increase the effect induced by a first pairing procedure, especially in 4–8 1-week-old kittens (Fig. 2).

Ocular dominance changes generally corresponded to differential modifications of the amplitude of the response for the two test stimuli (75% of the cases), and more rarely to a complex reorganization of the temporal structure of the peristimulus time histograms (PSTHs) (but see effects of the first pairing P_i on response to S^+ in Fig. 2). There were sometimes changes in the size of the receptive field, unmasking initially subliminal zones, but no response appeared *de novo*. In particular, four cells initially classified as being activated monocularly, remained driven exclusively by the same eye, in spite of repeated periods of imposed increase in firing concomitant with visual stimulation of the presumptive receptive field of the silent eye.

Similar findings were obtained using the orientation protocol (P_{ii}). Forty-three per cent of the paired cells showed a significant change in their relative preference between two stimuli shown through the same eye, but differing in orientation (range 24°–90°). In 94% of cases this shift was in favour of the stimulus paired with the 'high' level of activity. Changes appeared to be largest in kittens deprived of vision (Figs 3 and 4). A new orientation preference could be established in cells which were initially weakly biased or non-selective to orientation (Fig. 4). Modifications in orientation tuning studied on 45 cells showed that responsiveness was in most cases decreased for orientations close to S^- and increased for orientations close to S^+ . Shifts of preferred orientation of up to 90° could be obtained in normally reared or deprived kittens up to 12 weeks of age, whereas the largest change observed in adult was a shift of 30° (data not shown). Twenty-five per cent of modified cells maintained their orientation preferences, but adapted their response by significant changes in the polar asymmetry of the tuning profile. As in the case of ocular dominance, the pairing procedure does not induce *de novo* responses. One likely reason for a larger amplitude of effects described for kittens compared with the adult, could be the initially lower level of orientation selectivity and the broader spectrum of orientations through which neurones are activated in younger animals.

These experimentally induced changes in ocular dominance and orientation preference were never observed spontaneously⁶ and could not be explained by global modifications in levels of excitability and spontaneous activity, which were occasionally observed following pairing and pseudo-pairing (27% increase and 22% decrease). The minimum number of paired stimulations required to produce a significant effect was eighteen (1st pairing, Fig. 2), but for most cells 80 pairings were performed before a new control. Most changes appeared associative, since only 3% of 32 cells tested modified their relative preference after iontophoretic action uncorrelated with visual stimulation (that is, after pseudo-pairing). Modification was generally activity dependent, since changes were observed in 50% of the cells where a significant control of activity was imposed during pairing, and in only 8% of the cells for which iontophoresis did not affect visual response levels during pairing.

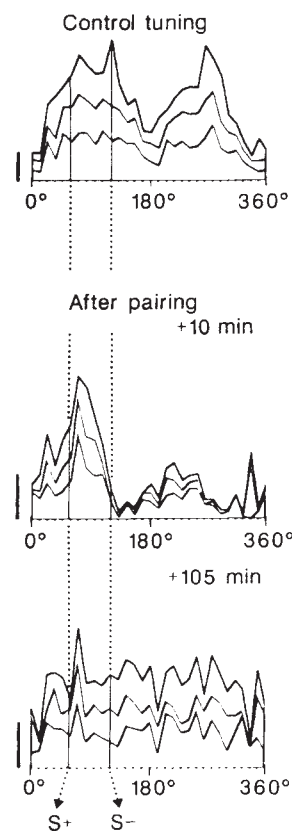


Fig. 4 Orientation selectivity change. Recordings were made from a cell in area 17 of a 5-week-old visually-deprived kitten. Orientation tuning curves were established and repeated three times over a 360° range (with steps of 12°) by randomized exploration of moving bars across the receptive field. Summed orientation tuning curves are shown before pairing (control tuning) and 10 min and 105 min after pairing. The pairing procedure consisted of 40 presentations of S^+ (60°) paired with a +4 nA current, interleaved in blocks of four with 40 presentations of S^- (120°) paired with a -5.6 nA current. The change in relative preference between the two test stimuli in favour of S^+ (iterative protocol), measured during the first hour following pairing (K.S., $P < 0.002$), was correlated with the modification shown in the orientation tuning profile established at the same time (middle row). Note the increased selectivity for S^+ and the loss of response for S^- . This cell, which was initially non-selective to orientation (responded to every direction of stimulation and was activated equally by a bar and a spot of light), acquired after pairing an orientation preference tuned closely with that of S^+ . This new functional preference extinguished within 2 hours. Calibration, 20 a.p.

These results are consistent with the hypothesis that changes in covariance levels between pre- and postsynaptic activities control the efficiency of transmission of already existing synapses. The finding that separate zones in the receptive field can be modified differentially (see second pairing in Fig. 2, where increase in visual response is restricted to the sub-zone which has been 'reinforced') is in favour of selective changes in synaptic efficiency. An increase in the efficiency of synapses which are active during S^+ pairing is predicted in hebbian schemes of plasticity, and recent reports from motor cortex^{18–20}, hippocampus^{21–23} and visual cortex²⁴ demonstrate that forced temporal correlation between electrical activation of afferent fibres and intracellular depolarization of the target neurone above a certain threshold, can induce an increase in the gain of excitatory synapses even in adult animals. By contrast, the decrease in the efficiency of synapses activated during S^- pairing, predicted by the covariance hypothesis and supported by our own data, is less well documented in the literature.

The study reported here is the first demonstration at the functional level, of changes analogous to those described classically during epigenesis of visual cortex^{5,6,11}. It is known that unilateral eye-lid closure following normal visual experience¹² or monocular vision in previously deprived kittens²⁵ results in a global shift of ocular dominance in favour of the eye remaining open. These rearing procedures could correspond to an S⁺ situation, where active or passive visuomotor interaction primes neuronal excitability in response to visual stimulation through the open eye^{25,26}. Furthermore, it has been shown recently that after monocular vision associated with blockade of postsynaptic activity (induced by intracortical injection of an agonist of GABA) most cells would respond only to the closed eye²⁷. This result is predicted by the covariance hypothesis, and is analogous to the effects we found after S⁻ pairing alone or interleaved with presentation of a neutral stimulus S⁰. Similar reasoning applies to the orientational protocol: our data demonstrate the capacity for certain cells to capture the orientation seen during a restricted visual exposure^{13,14}. However, it appears that cells can adapt their preference in favour of the imposed orientation, only if this latter initially evokes some response.

Finally, our study suggests that a hitherto unsuspectedly high level of plasticity is retained in the adult cortex. This favours the hypothesis that hebbian-like mechanisms operate both during development, and in adult learning^{4,8}. We suggest that, in the intact animal, extraretinal gating signals^{4,16,28}—instead of our pairing artifact—produce covariance changes through normal visuomotor experience during a postnatal critical period in kittens¹⁶, or during selective periods of learning in the adult^{28,29}. These would allow cortical neurons to undergo transition from a passive relay mode of transmission to an adaptive state reached only below or above certain levels of membrane potential. It still remains to be determined why there is an age-dependency in the expression of this experience-sensitivity in the intact, behaving animal⁸. In the adult animal, predominance of inhibitory intracortical networks and a lesser efficiency of the mechanisms responsible for the detection and transduction of these covariance changes^{24,30}, might reduce the probability of cortical synapses reaching the threshold for functional plasticity.

This work was supported by grants from the CNRS, MRT and FRM to Y. Frégnac. D. Shulz was supported by grants from Foundations Cino del Duca and Singer-Polignac. We thank Drs François Crépel, Henry Kennedy and Philip Hicks for helpful comments.

Received 7 March; accepted 8 April 1988.

1. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
2. Rauschecker, J. P. & Singer, W. *J. Physiol., Lond.* **310**, 215–239 (1981).
3. Bienenstock, E., Cooper, L. N. & Munro, P. *J. Neurosci.* **2**, 32–48 (1982).
4. Frégnac, Y. in *Imprinting and Cortical Plasticity* (eds Rauschecker, J. P. & Marler, P.), 221–266 (Wiley, New York, 1987).
5. Wiesel, T. N. *Nature* **299**, 583–591 (1982).
6. Frégnac, Y. & Imbert, M. *Physiol. Rev.* **64**, 325–434 (1984).
7. Sejnowski, T. J. *J. math. Biol.* **4**, 303–321 (1977).
8. Frégnac, Y. & Shulz, D. in *Seeing Contour and Colour* (ed. Kulikowski, J. J.) (Pergamon, Oxford, in the press).
9. Marr, D. *Proc. R. Soc. Lond. B*, **176**, 161–234 (1970).
10. Stent, G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 997–1001 (1973).
11. Sherman, S. M. & Spear, P. D. *Physiol. Rev.* **62**, 738–855 (1982).
12. Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **206**, 419–436 (1970).
13. Hirsch, H. V. B. & Spinelli, D. N. *Science* **168**, 869–871 (1970).
14. Blakemore, C. & Cooper, G. F. *Nature* **228**, 477–478 (1970).
15. Barlow, H. B. *Nature* **235**, 199–204 (1975).
16. Trotter, Y., Frégnac, Y. & Buisseret, P. *J. Neurophysiol.* **58**, 795–815 (1987).
17. Eccles, J. C. *The Physiology of Synapses* (Springer, Berlin, 1966).
18. Baranyi, A. & Féher, O. *Nature* **290**, 413–415 (1981).
19. Baranyi, A. & Szenté, M. *Brain Res.* **423**, 378–384 (1987).
20. Bindman, L. J., Meyer, T. & Pocket, S. J. *Physiol., Lond.* **386**, 90P (1987).
21. Wigström, H., Gustafsson, B., Huang, Y. Y. & Abraham, W. C. *Acta physiol. scand.* **126**, 317–318 (1986).
22. Kelson, S. R., Ganong, A. H. & Brown, T. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5326–5330 (1986).
23. Andersen, P. in *The Neural and Molecular Mechanisms of Learning* (eds Changeux, J. P. & Konishi, M.), 239–267 (Wiley, Chichester, 1987).
24. Ariola, A. & Singer, W. *Nature* **330**, 649–652 (1987).
25. Trotter, Y., Frégnac, Y. & Buisseret, P. C. *r. Acad. Sci. Paris D* **296**, 665–668 (1983).
26. Frégnac, Y. & Bienenstock, E. *Documenta Ophth. Proc. Ser.* **30**, 100–108 (1981).
27. Reiter, H. O. & Stryker, M. P. *Soc. Neurosci. Abstr.* **13**, 1241 (1987).
28. Kasamatsu, T., Watabe, K., Heggelund, P. & Schöller, E. *Neurosci. Res.* **2**, 365–386 (1985).
29. Heggelund, P., Imamura, K. & Kasamatsu, T. *Expl Brain Res.* **68**, 593–605 (1987).
30. Tsumoto, T., Hagihara, K., Sato, H. & Hata, Y. *Nature* **327**, 513–514 (1987).

Removal of phosphorylation sites from the β_2 -adrenergic receptor delays onset of agonist-promoted desensitization

Michel Bouvier, William P. Hausdorff, Antonio De Blasi, Brian F. O'Dowd, Brian K. Kobilka, Marc G. Caron & Robert J. Lefkowitz

Howard Hughes Medical Institute Laboratories, Departments of Physiology, Medicine and Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

Eukaryotic cells have evolved a variety of mechanisms for dampening their responsiveness to hormonal stimulation in the face of sustained activation. The mechanisms for such processes, collectively referred to as desensitization, often involve alterations in the properties and number of cell-surface hormone receptors^{1–3}. It has been speculated that phosphorylation–dephosphorylation reactions, which are known to regulate the catalytic activities of enzymes, also regulate the function of receptors⁴. Highly specific receptor kinases, such as rhodopsin kinase⁵ and β -adrenergic receptor kinase⁶, which show stimulus-dependent phosphorylation of receptors have been described. Direct evidence for a causal relationship between receptor phosphorylation and desensitization has been lacking however. Here we report that prevention of agonist-stimulated β_2 -adrenergic receptor (β_2 AR) phosphorylation by truncation of its serine and threonine-rich phosphate acceptor segment delays the onset of desensitization. We also show that selective replacement of these serine and threonine residues by alanine and glycine delays desensitization even further. These data provide the first direct evidence that one molecular mechanism of desensitization of G-protein-coupled receptors involves their agonist-induced phosphorylation.

Figure 1a depicts how G-protein coupled receptors, in this case the human β_2 AR, may be organized within the plasma membrane. Such proteins are thought to cross the membrane seven times^{7–9}. Light-dependent phosphorylation of a homologous G-protein coupled receptor, the visual pigment rhodopsin, by an enzyme termed rhodopsin kinase, occurs on multiple serine and threonine residues at its carboxyl terminus^{10–12}. A similar concentration of serine and threonine residues at the β_2 AR carboxyl terminus has been proposed as a target for agonist-promoted phosphorylation⁸.

To evaluate the role of phosphorylation in desensitization we constructed a mutant human β_2 AR complementary DNA encoding a protein truncated after amino-acid residue 365 (T-365). As shown in Fig. 1a, this mutant lacks most of the serine and threonine-rich carboxyl segment of the receptor. When expressed in Chinese hamster fibroblast CHW cells, both wild-type and mutant (T-365) β_2 AR bound the specific β AR ligand [¹²⁵I]-cyanopindolol (¹²⁵I-CYP) with high affinity ($K_D \approx 40$ pM) and appropriate β_2 AR specificity (data not shown). Moreover, photoaffinity labelling of both receptors in the CHW cells confirmed their expected mobility¹³ on SDS-PAGE (Fig. 1b).

To determine whether T-365 undergoes agonist-promoted phosphorylation, we equilibrated CHW cells expressing wild-type or T-365 β_2 AR with ³²P_i, exposed them to the β -agonist isoprenaline (2 μ M) for 15 min and then purified the β ARs after solubilization by affinity chromatography on alprenolol-Sepharose. The results are shown in Fig. 2. The levels of basal phosphorylation (apparent stoichiometries, see Fig. 2 legend) of the wild-type and mutant receptor were essentially identical. As previously shown in a variety of cell systems⁴, phosphorylation of the wild-type β_2 AR increased 2–3 fold when cells were exposed to the agonist isoprenaline (Fig. 2a). In contrast, no agonist-promoted increase in phosphorylation of the truncated receptor was observed at any time point investigated (15 min,

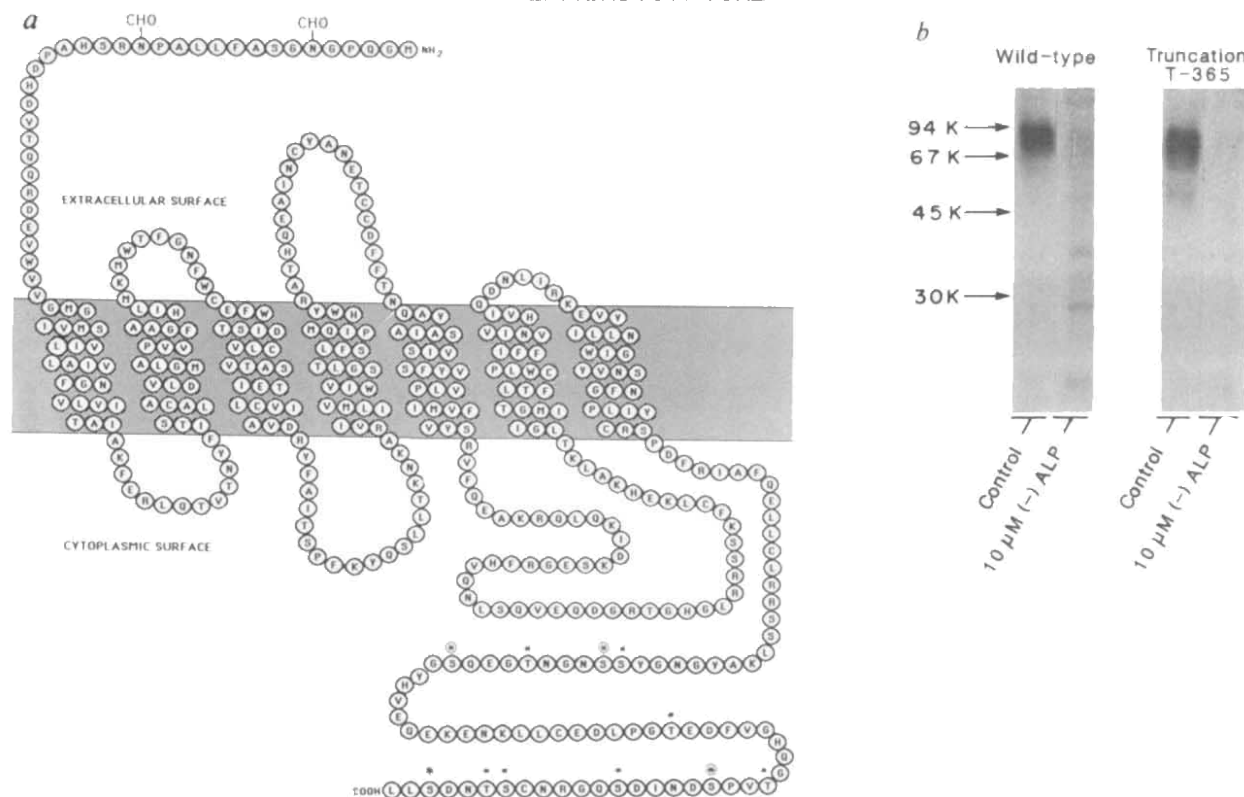


Fig. 1 *a*, Membrane topography of the wild-type and mutant human β_2 -adrenergic receptors predicted from hydropathicity analysis of the deduced amino-acid sequence¹⁹. The arrow indicates the site of truncation of the mutant T-365, and asterisks indicate substituted amino acids in S-351 (*, alanine, ⊗, glycine). *b*, Photoaffinity labelling of β_2 AR in membranes from CHW cells expressing wild-type (left panel) or T-365 (right panel) receptors.

Methods. To make the truncation mutant, wild-type β_2 AR cDNA was digested with *EcoRV* and religated to a blunt end adapter containing a termination codon and a *HindIII* site. The mutation was confirmed by loss of the *EcoRV* and appearance of the *HindIII* site. To create S-351, we used an oligonucleotide-directed mutagenesis system (Amersham). The human β_2 AR cDNA was cloned into the *EcoRI*-*HindIII* sites of the plasmid PTZ (Pharmacia) to enable preparation of single-stranded DNA template; mutagenesis was confirmed by sequencing. The wild-type and mutated human β_2 AR cDNAs, containing respectively 190 and 45 bp of 5'-untranslated sequence, were cloned into the eukaryotic expression vectors pKSV10 (Pharmacia) or pBC 12MI (ref. 20). The resulting plasmids were used with pSVNeo to cotransfect CHW cells by calcium phosphate precipitation²¹. Following selection, the clones expressing β_2 ARs were grown in Dulbecco's modified Eagle's medium (DMEM) plus 10% calf serum containing 150 mg ml⁻¹ geneticin. Several cell lines expressing the wild-type T-365 and S-351 β_2 ARs were generated and those expressing a similar number of receptors, (~ 1.5 pmol mg⁻¹ membrane protein) were selected. Essentially identical EC₅₀ concentrations (126 ± 13 nM, 134 ± 14 nM and 121 ± 16 nM respectively) for isoprenaline stimulation of adenylyl cyclase were observed for wild-type, T-365 and S-351 receptors. Characterization of the parental CHW cells and CHW cells expressing wild-type human β_2 AR has been published elsewhere¹³. For photoaffinity labelling, membranes were incubated in the dark with [¹²⁵I]iodocyanopindolol diazine (25 pM) in phosphate-buffered saline (PBS) containing 5 mM EDTA, in the presence or absence of 10 μ M alprenolol to define specific labelling, for 3 h at 25 °C. Following three washes with PBS-EDTA the membranes were resuspended in 1 ml of PBS-EDTA and UV-irradiated with a Hanovia 450 W medium pressure mercury lamp for 5 min. Autoradiograms of samples run on 12% SDS-PAGE (ref. 22) are shown. Results are representative of two separate experiments.

Fig. 2*b*; 5 min, 2 h, 5 h, data not shown).

These results indicated that truncation of the receptor prevents its agonist-induced phosphorylation; we therefore sought to determine the effect on agonist-promoted desensitization. We exposed wild-type and mutant receptor-containing CHW cells to 2 μ M isoprenaline for 2–180 min and examined the effect on subsequent isoprenaline-stimulated adenylyl cyclase activity in a cell membrane fraction and the extent to which receptors were internalized. When the wild-type receptors were exposed to agonist, desensitization was apparent even at the earliest time point studied (2 min). This continued to evolve at 10, 60 and 180 min and involved both a decrease in the maximum isoprenaline-stimulated activity, and a rightward shift in the agonist concentration response curve (Fig. 3*a*). No reduction in basal, NaF- and prostaglandin E₁-stimulated adenylyl cyclase activities was observed at 10 min (data not shown), indicating that the early desensitization is agonist-specific (homologous). Similar early desensitization was observed at 2, 5 and 10 min using the human adenocarcinoma cell line A431 which naturally expresses β_2 AR (data not shown).

In contrast, no such early desensitization (2 and 10 min) was observed in experiments with the truncated receptor (Fig. 3*b*). By 60 min, however, cells containing the mutant receptor had become desensitized to the same extent as the cells containing the wild-type receptor (Fig. 3*b*). This delay in the onset of desensitization was observed in each of four different clonal lines of T-365.

Agonist-induced sequestration (internalization) of the β_2 AR, however, as assessed by the binding of the hydrophilic ligand [³H]CGP12177 to surface receptors in whole cells, appeared to be amplified in the cells expressing the truncated β_2 AR (Fig. 4). The absence of desensitization at 10 min, despite sequestration of $\sim 30\%$ of the receptors in the T-365 cells, suggests that the full contingent of β_2 ARs is not required to maximally stimulate adenylyl cyclase in cells containing such a large number of receptors.

Truncation of the serine and threonine-rich carboxyl terminal segment removes $\sim 12\%$ of the β_2 AR mass, and might lead to conformational changes in the receptor protein unrelated to the absence of phosphorylation at the carboxyl terminus. Thus, we

Fig. 2 Whole-cell phosphorylation of the β_2 AR in CHW cells expressing wild-type or T-365 β_2 AR. Right: autoradiograms of affinity-purified β_2 AR solubilized from CHW cells expressing wild-type (a) or T-365 (b) β_2 AR that were prelabelled with carrier-free 32 P_i and incubated in the presence or absence of isoprenaline (2 μ M) for 15 min. The amount of 32 P in the excised bands was quantified by liquid scintillation spectrometry (wild-type 3,522 c.p.m. pmol⁻¹ in control and 8,645 c.p.m. pmol⁻¹ in isoprenaline exposed cells; T365, 3,285 c.p.m. pmol⁻¹ in control and 2,934 c.p.m. pmol⁻¹ in isoprenaline exposed cells). Left: densitometric scans of the autoradiograms corrected for the amount of β_2 AR (as determined by 125 I-CYP binding) loaded in each lane (a, lanes 1 and 2, 25 fmol each; b, lanes 1 and 2, 40 fmol each). This figure is representative of four experiments.

Methods. Nearly confluent cells expressing either wild-type or T-365 β_2 AR were detached from flasks by treatment with collagenase (1 mg ml⁻¹) containing soybean trypsin inhibitor (0.05 mg ml⁻¹) for 60 min at 37°. Following a thorough washing, the cells were resuspended in phosphate-free DMEM and incubated with carrier-free 32 P_i (0.25 mCi ml⁻¹ of cell suspension) at 37° for 60 min to allow labelling of the cellular ATP pool. The cells were then exposed to isoprenaline (2 μ M) for 15 min and the incubation was terminated by centrifugation at 200g. After washing with ice-cold phosphate-free DMEM, the cells were disrupted and plasma membranes prepared as described¹³. The β_2 AR was then solubilized with digitonin (2%) and purified by alprenolol-Sepharose affinity chromatography as described²³.

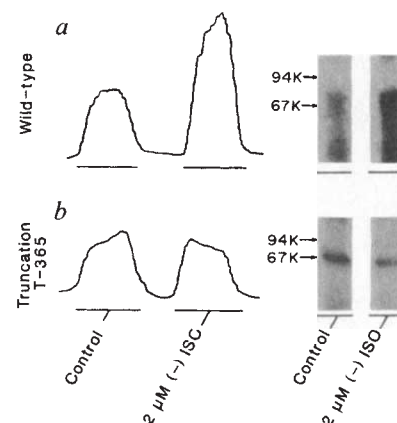
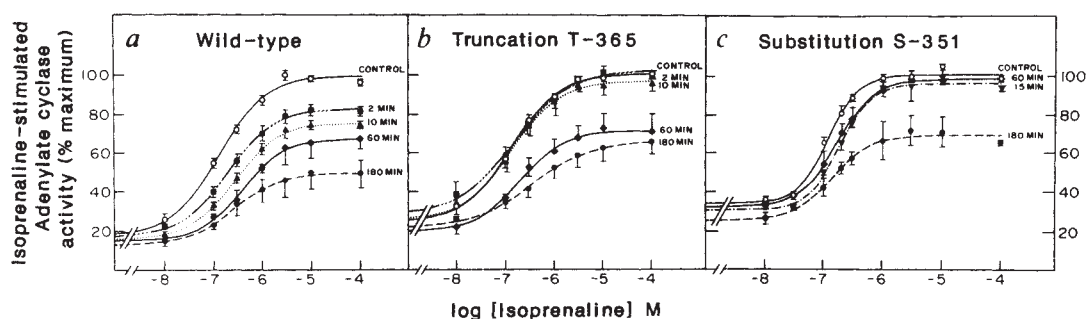


Fig. 3 Effects of isoprenaline pretreatment of β_2 AR-containing CHW cells on isoprenaline-stimulated adenylyl cyclase activity. CHW cells expressing the wild type (a) the T-365 (b) or the S-351 β_2 AR (c) were incubated in the absence (control) or presence of isoprenaline (2 μ M) for 2–180 min, and the ability of increasing concentrations



of isoprenaline to stimulate adenylyl cyclase in membranes derived from these cells was subsequently tested. The data shown are the mean of the following number of experiments: (control: wild-type $n=13$; T-365 $n=10$; S-351 $n=5$). (2 min: wild-type $n=8$; T-365 $n=5$; 10 min: wild-type $n=10$; T-365 $n=8$; 15 min: S-351 $n=6$; 60 min: wild-type $n=5$; T-365 $n=3$; S-351 $n=5$; 180 min: wild-type $n=4$; T-365 $n=2$; S-351 $n=3$.) The absolute (pmol min⁻¹ mg⁻¹) basal and maximal isoprenaline-stimulated adenylyl cyclase activity were: (wild-type: basal activity: control, 9.3 ± 1.1 ; 2 min, 10.2 ± 1.2 ; 10 min, 9.2 ± 1.2 ; 60 min, 6.6 ± 1.2 ; 180 min, 6.1 ± 1.8 . Maximum isoprenaline stimulation: 54.0 ± 6.8) (T-365: basal activity: control, 8.7 ± 1.5 ; 2 min, 12.1 ± 1.5 ; 10 min, 9.2 ± 1.2 ; 60 min, 4.8 ± 1.4 ; 180 min, 7.9 ± 1.2 . Maximum isoprenaline stimulation: 29.0 ± 4.3) (S-351: basal activity: control, 23 ± 2.3 ; 15 min, 22.3 ± 3.6 ; 60 min, 21.5 ± 1.6 ; 180 min, 12.8 ± 1.8 . Maximum isoprenaline stimulation: 68.2 ± 8.9). Values are expressed as the mean $\pm$ s.e.m..

Methods. Isoprenaline (2 μ M) was added to the cells with fresh DMEM containing 10% fetal calf serum and the cells were incubated at 37°C for the indicated period of time. The cells were then detached by rapid scraping and plasma membranes were prepared as described¹³. Adenylyl cyclase activity was determined by the method of Salomon *et al.*²⁴ and expressed as % $\pm$ s.e.m. of the maximum stimulation in the control.

examined the pattern of agonist-induced desensitization of a β_2 AR mutant with a full length carboxyl terminus, but in which the potential phosphate acceptor serines and threonines (a total of 11) had been mutated to alanine or glycine (S-351, Fig. 1a). As shown in Fig. 3c, in cells containing such a mutant receptor no desensitization was observed following exposure to isoprenaline for 15 min and 60 min. By 180 min, however, these cells did show significant desensitization. Agonist-induced sequestration of S-351 was identical to that observed in the cells expressing the wild-type receptor (data not shown).

The data described here represent the most direct evidence yet developed for a causal relationship between agonist-promoted phosphorylation of the β_2 AR and its desensitization. In addition, these results indicate that the serine and threonine-rich carboxyl terminus of β_2 AR is a major site of agonist-promoted phosphorylation in whole cells. This segment of the receptor has been proposed as the target of the β -adrenergic receptor kinase *in vitro*⁸ and thus this enzyme is strongly implicated as being responsible for the early agonist-promoted phosphorylation of the receptor in whole cells.

Our data further show, in agreement with several other lines of evidence^{14,15}, that the early phase of homologous desensitization can be distinguished from the sequestration of the receptor. Indeed, in cells expressing T-365 or S-351 in which the onset of desensitization is significantly delayed, the sequestration was either greater than, or identical to, that observed in cells express-

ing the wild-type receptor. Agonist-induced phosphorylation of the receptor at the carboxyl terminus, therefore, does not appear to be essential for sequestration. The biological significance of the amplified sequestration in T-365, however, which is in contrast with a previous report¹⁶ that a very similar mutant sequestered normally when expressed in mouse L cells, remains to be determined.

Desensitization is delayed rather than abolished in the mutants lacking the carboxy-terminal phosphorylation sites. Thus, additional mechanisms such as receptor sequestration, heterologous desensitization or binding of a putative arrestin-like molecule to the receptor¹⁷ may contribute to the more slowly evolving adaptation. Alternatively, phosphorylation may modulate the rate rather than the extent of desensitization. Such an effect has been proposed for the nicotinic cholinergic receptor where phosphorylation of the receptor-channel by the cAMP-dependent protein kinase *in vitro* enhances the rate at which the channel desensitizes¹⁸.

Parallels can also be drawn between our data and the recently suggested involvement of the serine and threonine-rich carboxyl terminus in the adaptation processes of yet another putative member of the G-protein-coupled receptor family. Mutants of the yeast α -mating factor receptor, lacking the carboxyl terminus, display a supersensitive phenotype with regard to the cell division arrest normally induced by the α -factor (Remeke, J. and Thorner, J., manuscript in preparation). Such truncation

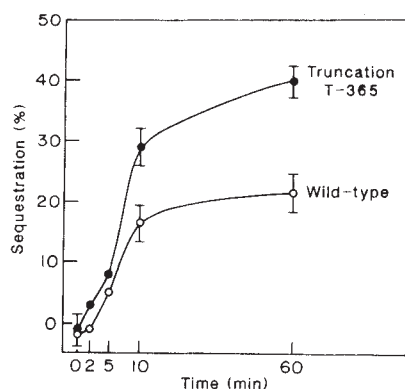


Fig. 4 Isoprenaline-induced sequestration of β_2 AR in CHW cells expressing wild-type or T-365 β_2 AR. Cells were incubated with isoprenaline ($2 \mu\text{M}$) at 37° for the indicated times, extensively washed and detached by rapid scraping with cold PBS as described¹³. A trypan-blue exclusion test showed that $>95\%$ of the cells were intact following scraping. Two different aliquots from the same flask were used for measuring total and surface receptors. Total receptors were determined by [^{125}I]iodopindolol ($\sim 150 \text{ pM}$) binding at 37° for 60 min. Surface receptors were determined by [^3H]-CGP 12177 ($\sim 3 \text{ nM}$) binding at 4°C for 18 h²⁵. Nonspecific binding was defined in the presence of $1 \mu\text{M}$ (-) propranolol. The % of sequestration was calculated as: $[1 - (\text{surface receptors} / \text{total receptors})] \times 100$. The total number of cell receptors did not change over the time course shown. Following 3 h isoprenaline treatment (not shown) a decrease in total cell receptors of $\sim 25\%$ was observed in both wild-type and T-365 containing cells. The data with error bars represent the mean $\pm$ s.e.m. of four separate experiments. The data without error bars represent the mean of two separate experiments.

mutants are thus resistant to the 'desensitization' ordinarily induced by the mating factor. Such analogies suggest that receptor phosphorylation on carboxy-terminal cytoplasmic domains, probably mediated by receptor-specific kinases, may be of very broad regulatory significance.

We thank Grace Irons, Tong Sun Kobilka and Mark Leader for technical support, Dr. Carolyn MacGregor for help in constructing S-351, Mary Holben for preparation of the manuscript and Dr. Jeffrey Benovic for critical reading of the manuscript. This work was supported in part by NIH. M.B. holds a Centennial fellowship from the MRC of Canada, A.D.B. is a fellow in the program of advanced training established by EEC and Regione Lombardia, Italy.

Received 22 January; accepted 15 April 1988.

- Harden, T. K. *Pharmac. Rev.* **35**, 5-23 (1983).
- Clark, R. B. *Adv. Cyclic Nucleotide Res.* **20**, 155-209 (1986).
- Sibley, D. R. & Lefkowitz, R. J. *Nature* **317**, 124-129 (1985).
- Sibley, D. R., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *Cell* **48**, 913-922 (1987).
- Kuhn, H. & Dreyer, W. J. *FEBS Lett.* **20**, 1-6 (1972).
- Benovic, J. L., Strasser, R. H., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2797-2801 (1986).
- Dixon, R. A. F. *et al. Nature* **321**, 75-79 (1986).
- Dohlman, H. G., Bouvier, M., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **262**, 14282-14288 (1987).
- Henderson, R. & Umwin, P. N. T. *Nature* **257**, 28-32 (1975).
- Hargrave, P. A. *et al. Neurochem. Int.* **1**, 231-244 (1980).
- Thompson, P. & Findlay, J. B. C. *Biochem. J.* **220**, 773-780 (1984).
- Miller, J. L. & Dratz, E. A. *Vision Res.* **24**, 1509-1521 (1984).
- Bouvier, M. *et al. Molec. Pharmac.* **33**, 133-139 (1988).
- Homburger, V. *et al. J. biol. Chem.* **255**, 10436-10444 (1980).
- Waldo, G. L., Northup, A. J. K., Perkins, J. P. & Harden, T. K. *J. biol. Chem.* **258**, 13900-13908 (1983).
- Strader, C. D. *et al. Cell* **49**, 855-863 (1987).
- Benovic, J. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **24**, 8879-8882 (1987).
- Huganir, R. L., Delcour, A. H., Greengard, P. & Hess, G. P. *Nature* **321**, 774-776 (1986).
- Kobilka, B. K. *et al. J. biol. Chem.* **262**, 7321-7327 (1987).
- Cullen, B. *Meth. Enzym.* **152**, 684 (1987).
- Gorman, C. in *DNA cloning, A Practical Approach* Vol. II. (ed. Glover, D. M.) 143-290 (IRL Press, Oxford, 1985).
- Laemmli, U. K. *Nature* **227**, 680-686 (1970).
- Shorr, R. G. L., Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **256**, 5820-5826 (1981).
- Salomon, Y., Londos, C. & Rodbell, M. *Analyt. Biochem.* **58**, 541-548 (1974).
- DeBlasi, A., Lipariti, M., Motulsky, S., Insel, P. A. & Fratelli, M. *J. clin. Endocr. Metab.* **61**, 1081-1088 (1985).

Conformational changes associated with ion permeation in L-type calcium channels

Daniela Pietrobon, Blaise Prod'homme & Peter Hess

Department of Cellular and Molecular Physiology and Program in Neuroscience, Harvard Medical School, Boston, Massachusetts 02115, USA

The mechanism by which ions deliver their message to effector proteins involves a change in the protein conformation which is induced by the specific interaction of the ion with its binding site on the protein. In the case of an ion-channel protein, conformational changes induced by permeant ions and the consequences for channel function have received little attention. Here we report that binding of permeant cations to an intra-channel binding site of the dihydropyridine (DHP)-sensitive (L-type) Ca^{2+} channel leads to a conformational change which destabilizes the protonated state of a group on the external channel surface, and can shift its apparent pK value by more than 2 pH units. The lifetime of the protonated state correlates with the occupancy of an intra-channel binding site by the permeant cation. The demonstration of such conformational changes in a channel protein induced by the permeant ion has important implications for realistic models of the mechanism of ion permeation.

We have recently demonstrated that protonation of a site located at the external surface of the L-type Ca channel greatly reduces the channel conductance when Na^+ ions carry the current in the absence of divalent ions¹. When 110 mM Ba^{2+} is the charge carrier, however, the conductance of L-type Ca channels is little affected between values of external pH (pH_0) of 6-9, suggesting a correlation between the protonation reaction and the species of permeant ion. Figure 1 shows that this is indeed the case, and that the protonation kinetics are vastly different even when three monovalent charge carriers are compared. The top traces in Fig. 1 show the typical elementary currents of L-type Ca channels in the presence of the DHP Ca channel agonist (+)-(S)-202-791 (refs 2, 3) when Cs^+ is the charge carrier. The opening of the channel is followed by transitions between two different conducting levels. Our interpretation¹ that these transitions reflect binding and unbinding of individual protons at a single protonation site is supported by the finding that the mean time the channel spends at the high conductance level decreases linearly with increasing H^+ concentration (ref. 1 and unpublished data). A similar pattern of single channel currents is seen when K^+ is the charge carrier (Fig. 1, middle traces), but the lifetime of the low conductance state is shorter than with Cs^+ . With Na^+ as the charge carrier (Fig. 1, bottom traces), transitions between the two levels are so fast that they can no longer be resolved at the bandwidth of our recording system (5 kHz) and become apparent as large open channel noise¹. The pH_0 -values for the traces in Fig. 1 were chosen such that with each permeant ion the open channel spends approximately equal time in the protonated and unprotonated state, as seen most directly from the similar peaks at the two levels in the amplitude histogram for each ion (Fig. 1). Therefore the apparent pK of the site is close to the pH_0 in each ionic condition shown, and it is clear that the pK shifts by more than 1 pH unit as Na^+ replaces Cs^+ as the permeant ion.

A plot of the protonation and deprotonation rate constants with each of the three ions (Fig. 2a) shows that the shift of the pK value results mainly from differences in the deprotonation rate constant ($k_{\text{off,H}}$). The average time the channels spend at the low conductance level decreases from 480 μsec in the presence of Cs^+ to 60 μsec when Na^+ is the charge carrier.

Because the L-type Ca channel distinguishes between permeant ions primarily on the basis of the ion's affinity to intra-

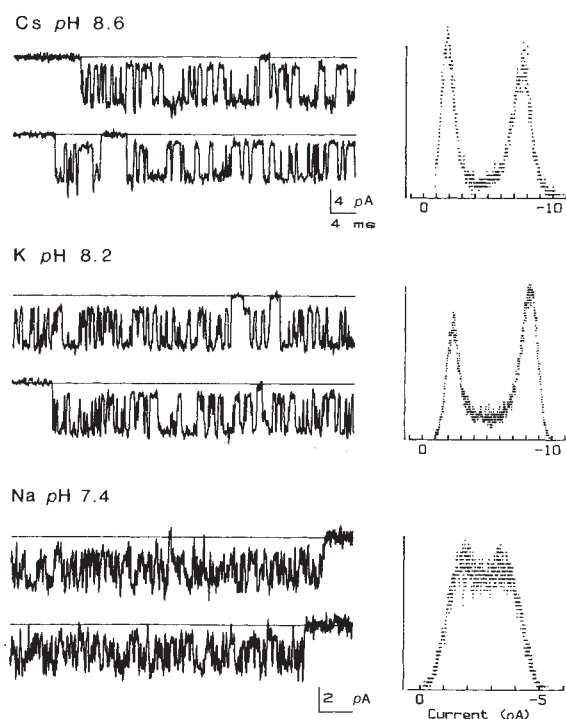


Fig. 1 Cell-attached patch recordings (left panels) and amplitude histograms (right panels) of single L-type Ca^{2+} channel currents carried by Cs^+ at pH 8.6 (top), K^+ at pH 8.2 (middle) or Na^+ at pH 7.4 (bottom). Level of closed channel is indicated by solid line in current traces and marked as 0 in amplitude histograms. Current segments of closed channels were not included in amplitude histograms, which therefore represent only current fluctuations between the two conducting levels. L-type Ca channels were activated by depolarizations from a holding potential of -90 to -110 mV and channel activity was recorded either during the activating pulse (at -40 mV with Na^+ as the charge carrier) or as unitary tail currents following repolarization to varying test potentials (-70 mV and -90 mV with K^+ or Cs^+ respectively). In these voltage protocols, which are dictated by the voltage dependence of L-type Ca channel gating in the absence of divalent ions^{1,10} and the negative reversal potentials when K^+ or Cs^+ are the charge carriers (see Fig. 2), we took advantage of the absence of voltage dependence of the protonation kinetics¹.

Methods. Cell-attached patch clamp recordings²⁰ were obtained from freshly dissociated guinea pig ventricular cells¹⁰, or undifferentiated pheochromocytoma (PC-12) cells cultured with standard techniques. The two preparations gave identical results. Patch pipettes contained 160 mM CsCl, KCl or NaCl (from top to bottom) and 5 mM EDTA. The pH buffer was TAPS (pK, 8.4) for the Cs^+ and K^+ solutions and HEPES (pK, 7.55) for the Na^+ solution. The cell membrane-potential outside the patch was zeroed¹⁰ by an external solution containing 140 mM potassium aspartate, 10 mM EGTA, 10 mM HEPES, 20 mM MgCl_2 . The DHP Ca^{2+} channel agonist (+)-(S)-202-791 (gift from Dr Hof, Sandoz Co.) was added at $1 \mu\text{M}$ to the external solution to promote long-lasting channel openings^{1,3}. The currents were digitized at 25 kHz and filtered at 5 kHz (-3dB , 8-pole Bessel filter). Temperature 20 – 22°C .

channel binding sites^{4–9}, we explored the possibility that the value of $k_{\text{off,H}}$ with each permeant ion was related to the ion's binding affinity for a permeation site. We judged the relative affinity of the channel for each ion by its conductance and zero-current potential (Fig. 2b). Channel conductance increases in the order $\text{Na}^+ < \text{K}^+ < \text{Cs}^+$, but as the conductance increases, the zero-current potential shifts towards more negative values. This combination of conductance and zero-current potential is the expected finding for selectivity by affinity^{9,10}: the ion which binds most tightly (in this case Na^+) has the lowest mobility (lowest conductance), but the most positive zero current potential, because it competes most effectively for channel occupancy

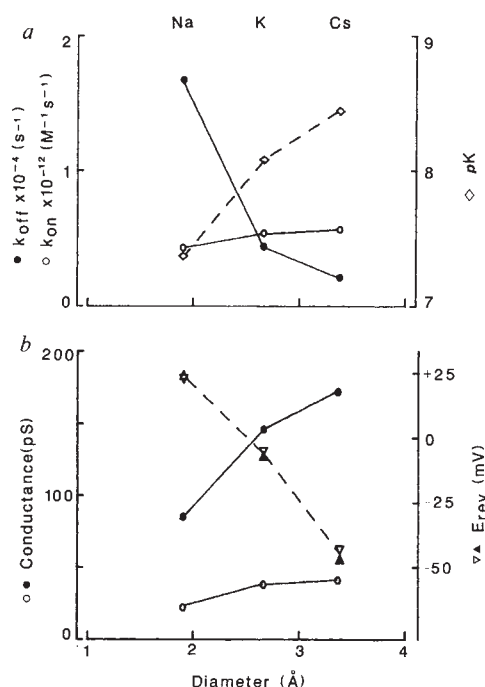


Fig. 2 Protonation kinetics and permeation parameters with Na^+ , K^+ or Cs^+ as the charge carrier. **A**, protonation rate constants (k_{on} , open circles), deprotonation rate constants (k_{off} , filled circles) and pK values (diamonds) with 160 mM Na^+ , K^+ , or Cs^+ in the pipette are plotted as a function of the diameter of the unhydrated ion. **B**, Unitary conductances (circles) and zero current potentials (triangles) of the unprotonated (filled symbols) and protonated channel (open symbols).

Methods. Protonation and deprotonation rate constants of K^+ and Cs^+ currents were obtained at pH 8.2 by kinetic analysis of the fluctuations between the two conductance levels¹. Histograms of the time intervals spent at the high (unprotonated) or low (protonated) conductance levels were fitted with a single exponential (nonlinear χ^2 minimization). Second-order protonation rate constants were obtained by division by the H^+ concentration in the pipette. The values for k_{on} , k_{off} and pK with Na^+ at pH 7.0–7.4 are those reported in ref. 1, Table 1, and are shown here for ease of comparison. The conductances for K^+ and Cs^+ were measured at pH 8.4–8.6. The conductances for Na^+ were obtained at pH 9 for the unprotonated channel and at pH 6 for the protonated channel. All current-voltage relations were well fitted by straight lines over the entire voltage range tested, and the zero-current potentials were extrapolated (linear regression) and corrected for liquid junction potentials. All symbols are means of 3–9 determinations. Standard errors of the mean are not shown because they were always smaller than the size of the symbols.

with the internal K^+ present in these cell-attached recordings. The affinity thus increases in the order $\text{Cs}^+ < \text{K}^+ < \text{Na}^+$, the same order in which Cs^+ , K^+ and Na^+ destabilize the protonated state. The protonated channel maintains its selectivity among monovalent cations as shown by the zero current potentials which are not measurably different for the protonated and unprotonated channel (Fig. 2b).

The conclusion of a direct link between channel occupancy and the deprotonation rate constant is further strengthened by the finding that $k_{\text{off,H}}$ increases when the K^+ concentration is increased from 60 to 160 mM (Fig. 3).

Ca^{2+} and Ba^{2+} ions bind much more strongly to the intra-pore sites than monovalent ions^{7,8}. At micromolar external Ca^{2+} concentrations, the dwell time of an individual Ca^{2+} ion in the channel is long enough to block flow of monovalent ions for a measurable time¹¹. The resulting elementary current traces at two Ca^{2+} concentrations are shown in Fig. 4a and compared with a control trace with K^+ in the absence of Ca^{2+} . As expected, in the presence of Ca^{2+} the long channel openings are now

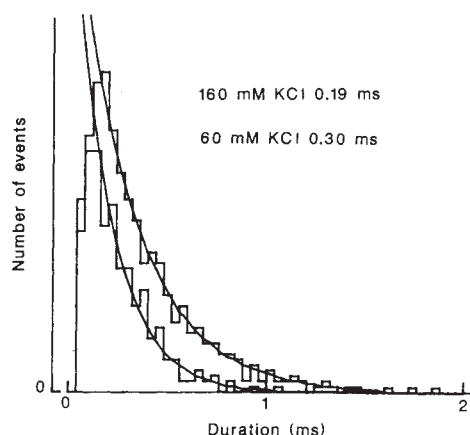


Fig. 3 The deprotonation rate constant increases with increasing concentration of permeant ion. Superimposed distributions of durations of the protonated state with 60 mM and 160 mM K^+ in the pipette at pH 8.4. The solid lines are χ^2 -minimizations of single exponentials with time constants of 0.3 and 0.19 ms for 60 and 160 mM K^+ respectively. The ordinates are scaled such that the number of events extrapolated by the fitting routine for time zero is equal for both distributions. The first three (160 mM K^+) or four (60 mM K^+) bins are excluded from the fitting procedure. To keep osmolality and ionic strength constant, 100 mM choline chloride was added to the pipette solution containing 60 mM K^+ .

interrupted by discrete transitions to the fully closed level, which we interpret as individual sojourns of a Ca^{2+} ion in the channel pore (see ref. 11 and legend to Fig. 4). As long as the Ca^{2+} ion is bound in the pore, we cannot judge its effect on the protonation kinetics. Our hypothesis that occupancy of the binding site destabilizes the low conductance (protonated) state, however, predicts that immediately following departure of the Ca^{2+} ion the channel should be in its high conductance (deprotonated) state. Inspection of the traces with Ca^{2+} in Fig. 4a shows that indeed almost all Ca^{2+} -unblocking transitions occur from the zero-current level to the fully open level corresponding to the deprotonated state. This is confirmed in a more rigorous way in Fig. 4b, where we display the time course of the mean current just preceding entry of a Ca^{2+} ion and that just following Ca^{2+} exit, obtained by lining up and averaging several hundred individual blocking and unblocking events. The current following the departure of the Ca^{2+} ion reaches an initial peak from which it relaxes back to the steady state value halfway between the two conductance levels, reflecting the fact that in the absence of Ca^{2+} at pH 8.2 the open channel spends approximately equal time at the protonated and unprotonated levels (Fig. 1, Fig. 4a, top trace). This relaxation shows unequivocally that binding of a Ca^{2+} ion to a site within the conduction pathway has biased the protonation equilibrium towards the unprotonated state. The site to which Ca^{2+} binds must be located within the pore because (1) the presence of a Ca^{2+} ion completely blocks current flow and (2) the departure rate constant is speeded up by hyperpolarization indicating that the exit occurs towards the inside of the channel (ref. 11 and unpublished data). Thus the experiment in Fig. 4 directly confirms the hypothesis that it is the occupation of an intra-pore binding site which increases $k_{off,H}$ and makes it very likely that the effects of the monovalent cations on $k_{off,H}$ are also mediated via the intra-channel site.

In contrast to the initial relaxation of the averaged current following Ca^{2+} exit from the pore, little deviation from the steady state level is seen in the averaged current preceding a Ca^{2+} blocking event (Fig. 4b). This means that Ca^{2+} entry occurs with roughly equal rate into the protonated and unprotonated channel and lets us conclude that the protonation site is far enough from the permeation pathway that an entering ion does not experience the different local potential associated with the protonated or unprotonated site. Thus, the different conduct-

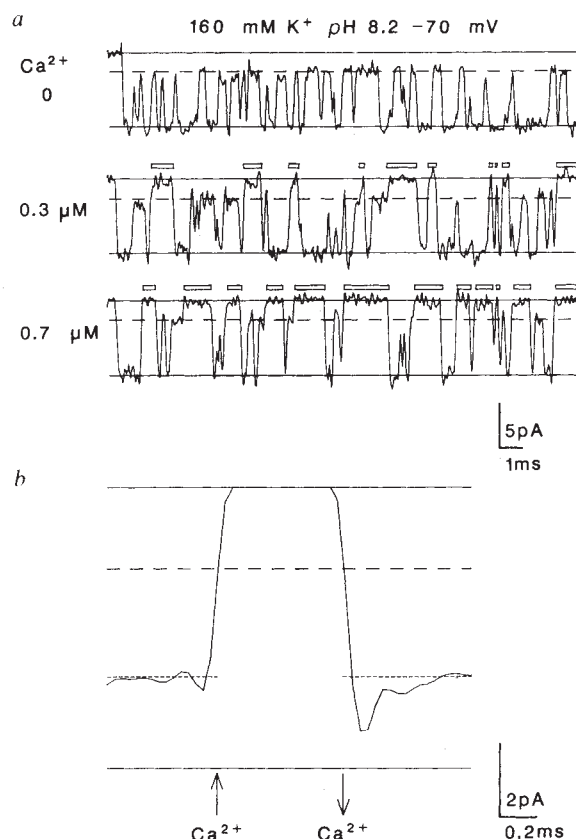


Fig. 4 Binding of a Ca^{2+} ion to an intra-channel site perturbs the protonation equilibrium. **A**, Current traces from three different patches obtained with 160 mM K^+ in the pipette at pH 8.2 and 0 (top trace), 0.3 μM (middle) and 0.7 μM Ca^{2+} . The pipette Ca^{2+} was buffered to the desired level with HEDTA ($K_{D, Ca}$, 0.33 μM at pH 8.2). The horizontal lines indicate the zero-current level and the two conductance levels of the protonated and unprotonated channel. In the presence of Ca^{2+} the records are interrupted by frequent transitions to the zero-current level. These sojourns of individual Ca^{2+} ions in the pore are indicated above the current records by horizontal bars. Kinetic analysis of the current traces at 0.7 μM Ca^{2+} yielded a mean dwell time of a Ca^{2+} ion of 0.70 msec and a mean interval between blocking events of 0.63 msec, corresponding to entry and exit rate constants of $2.2 \times 10^9 M^{-1} s^{-1}$ and $1430 sec^{-1}$ respectively. Patch potential $-70 mV$. **B**, Mean current just preceding the entry (left side) and just following right side the exit of a Ca^{2+} ion. Idealization of individual blocking and unblocking events was used to determine the time of alignment of the original events for the construction of the mean currents, which represent averages of several hundred events. The continuous horizontal lines indicate the three current levels of the unitary recordings. The interrupted dotted line indicates the level of average (steady state) proton block, which as expected at pH 8.2 and 160 mM K^+ is $\sim 50\%$ (see Figs 1, 2). The transient of the average current following the Ca^{2+} unblocking event is attenuated because of filtering and poor temporal synchronization of the alignment of events due to noise. When simulated events with obligatory first transitions to the fully open level were filtered, added to experimental background noise and analysed with the same programmes used to obtain the averages from the real traces, the resulting transient following the simulated unblocking event was very similar to that shown in Fig. 4b and also fell short of reaching the fully open level.

ances of the protonated and unprotonated channel cannot be explained by a protonation dependent surface potential at the channel entrance which changes the local concentration of permeant ions¹. A direct competition between protons and permeant cations for the same site can be excluded as well, as in this case a Ca^{2+} ion would only be able to enter the unprotonated site. Further arguments against competition of the two ions for a

common site include the finding that protons do not block the permeation pathway completely and that the effects of permeant ions on the protonation kinetics are mainly on $k_{\text{off,H}}$.

Thus we conclude that protons and permeant cations interact allosterically at different sites. Occupancy of the cation permeation site changes the protein conformation in a way that favours deprotonation of the proton site. In turn, protonation triggers a conformational change which reduces the channel conductance for monovalent ions with little change in selectivity.

The independence of the Ba^{2+} or Ca^{2+} conductance on pH between pH 6.5–9 can be explained by near-saturating channel occupancy which will destabilize the protonated state and strongly bias the equilibrium towards the high conductance (unprotonated) state.

The large changes in $k_{\text{off,H}}$, and the deviation of the apparent pK values from that of any unperturbed amino-acid side chain in a peptide chain, suggest short-range electrostatic interactions between the protonated group and its local environment (see ref. 12, for example). These electrostatic interactions appear to be very sensitive to changes in channel conformation, thus allowing us to probe the conformational changes associated with the coordination of permeant ions within the channel.

Whatever the exact molecular mechanism, our experiments emphasize that channels cannot be regarded as inert structures when the mechanism of ion permeation is considered. The conformational change induced by binding of the permeant ion is likely to be very rapid. At physiological millimolar Ca^{2+} concentrations, however, even short-lived conformational changes induced by a permeant ion could last long enough to be felt by the next ion entering the channel. In this case, compli-

cated permeation patterns like anomalous mole fraction effects^{7,8,13–15}, so far taken as evidence for multi-ion pores^{16,17}, would also be expected, even in channels with a single intra-channel binding site^{18,19}. Conformational changes, when specifically induced by a particular ion, could thus contribute to the selectivity of the channel and would have to be included in realistic models for channel permeation.

We thank Dr Mark Plummer for help with tissue culturing. This work was supported by the USPHS, fellowships by EMBO and Fogarty (D.P.) and a fellowship of the Muscular Dystrophy Association (B.P.).

Received February 12; accepted 11 April 1988.

1. Prod'homme, B., Pietrobon, D. & Hess, P. *Nature* **329**, 243–246 (1987).
2. Hof, R. P., Ruegg, U. T., Hof, A. & Vogel, A. J. *Cardiovascular Pharmacol.* **7**, 689–693 (1985).
3. Kokubun, S., Prod'homme, B., Becker, C., Porzig, H. & Reuter, H. *Molec. Pharmacol.* **30**, 571–584 (1986).
4. Vereeke, J. & Carmeliet, E. E. *Pflügers Arch. ges. Physiol.* **322**, 73–82 (1971).
5. Hagiwara, S., Fukuda, J. & Eaton, D. C. *J. gen. Physiol.* **63**, 564–578 (1974).
6. Kostyuk, P. G., Mironov, S. L. & Shuba, Y. M. *J. Membrane Biol.* **76**, 83–93 (1983).
7. Hess, P. & Tsien, R. W. *Nature* **309**, 453–456 (1984).
8. Almers, W. & McCleskey, E. W. *J. Physiol., Lond.* **353**, 585–608 (1984).
9. Coronado, R. & Smith, J. S. *Biophys. J.* **51**, 497–502 (1987).
10. Hess, P., Lansman, J. B. & Tsien, R. W. *J. gen. Physiol.* **88**, 293–319 (1986).
11. Lansman, J. B., Hess, P. & Tsien, R. W. *J. gen. Physiol.* **88**, 321–347 (1986).
12. Warshel, A. *Nature* **330**, 15–16 (1987).
13. Eisenman, G., Sandblom, J. P. & Walker, J. L. Jr. *Science* **155**, 965–974 (1967).
14. Neher, E. *Biochim. biophys. Acta* **401**, 540–544 (1975).
15. Hagiwara, S., Miyazaki, S., Krasne, S. & Ciani, S. J. *J. gen. Physiol.* **70**, 269–281 (1977).
16. Urban, B. W., Hladky, S. B. & Haydon, D. A. *Fedn Proc.* **37**, 2628–2632 (1978).
17. Hille, B. & Schwarz, W. J. *J. gen. Physiol.* **72**, 409–442 (1978).
18. Ciani, S. *Biophys. J.* **46**, 249–252 (1984).
19. Lauger, P. *Biophys. J.* **47**, 581–591 (1985).
20. Hamill, O., Marty, A., Neher, E., Sakmann, B., & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).

Control of neuronal fate by the *Drosophila* segmentation gene *even-skipped*

Chris Q. Doe*, David Smouse* & Corey S. Goodman†

Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

The central nervous system (CNS) contains a remarkable diversity of cell types. The molecular basis for generating this neuronal diversity is poorly understood. Much is known, however, about the regulatory genes which control segmentation and segment identity during early *Drosophila* embryogenesis^{1,2}. Interestingly, most of the segmentation and homoeotic genes in *Drosophila*, as well as many of their vertebrate homologues, are expressed during the development of the nervous system (for example, ref. 3). Are these genes involved in specifying the identity of individual neurons during neurogenesis, just as they specify the identity of cells during segmentation? We previously described the CNS expression of the segmentation gene *fushi tarazu* (*ftz*) and showed that *ftz* CNS expression is involved in the determination of an identified neuron³. Here we show that another segmentation gene, *even-skipped* (*eve*), is expressed in a different but overlapping subset of neurons. Temperature-sensitive inactivation of the *eve* protein during neurogenesis alters the fate of two of these neurons. Our results indicate that the nuclear protein products of the *eve* and *ftz* segmentation genes are components of the mechanism controlling cell fate during neuronal development.

In insects, neurogenesis begins with two steps. Initially, cell interactions control the differentiation of a segmentally repro-

ducible pattern of neuronal stem cells (called neuroblasts) from a morphologically uniform sheet of neuroectodermal cells^{4–7}. The second step involves the generation of a characteristic family of neurons from each neuroblast^{3,8–11}. Neuronal identity is determined in part by birth order, or cell lineage, from the neuroblast^{6,12}. In *Drosophila*, a group of genes called the neurogenic genes which play a role in the cell interactions governing neuroblast versus epidermal fate have been identified^{13,14}. Genes involved in the determination of neuroblast progeny, however, have yet to be identified.

Expression of many segmentation genes has been observed in the developing CNS^{15–23}. Segmentation gene expression fulfills several criteria expected of genes involved in neuronal determination. First, they are expressed early in neurogenesis, at the contrast to their initial expression in broad regions or stripes during segmentation, at the time of neurogenesis they are expressed in a characteristic subset of neurons in every segment of the developing CNS. We recently showed that the CNS expression of the segmentation gene *ftz* is involved in the determination of the RP2 neuron³. In this paper, we identify several neurons that express the segmentation gene *eve*, and show that *eve* CNS expression is essential for the correct determination of at least two of these neurons, RP2 and aCC.

The *eve* protein contains a homeodomain and is localized to the nucleus^{20,21}. The gene products of *eve* are initially distributed in seven circumferential stripes at the blastoderm stage which is superseded by a 14 stripe expression pattern by early gastrulation^{20,21}. After gastrulation, at about the third hour of development, all striped expression disappears. The *eve* protein is first detected in the CNS at about hour 5 and then persists throughout embryogenesis²¹. At hour 12 *eve* is expressed in 16 neurons (out of a total of ~250) per hemisegment. Three of these neurons have previously been identified and characterized: these are the sibling aCC and pCC neurons, and the RP2 neuron^{9,24–26} (its sibling, the RP1 neuron, does not express *eve*).

It is impossible to analyse the CNS function of *eve* in embryos lacking all *eve* expression. Loss of early blastoderm stripe expression alters the fates of blastoderm cells²⁷, and much of

* Present address: Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA (C.Q.D.) and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA (D.S.).

† To whom correspondence should be addressed at Department of Biochemistry, University of California, Berkeley, California 94720, USA.

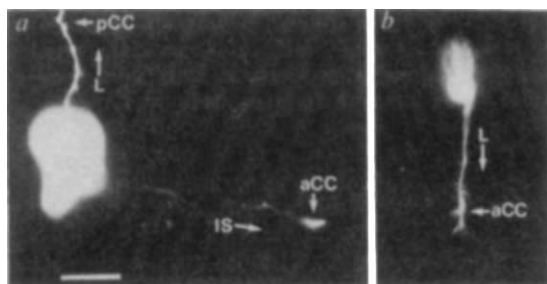


Fig. 1 Axon morphology of the aCC neuron in wild-type *Drosophila* embryo (a) and in the absence of *eve* CNS expression (b). a, Lucifer yellow injections²⁴ reveal the growth cones and axon morphology of the aCC (anterior of the two sibling neurons) and pCC neurons in an hour 10½ embryo. In wild-type embryos the aCC is a motoneuron, the growth cone of which extends laterally, pioneering the IS nerve. The pCC is an interneuron the growth cone of which extends anteriorly into the longitudinal connective. b, In this embryo lacking *eve* CNS function, the aCC growth cone extends posteriorly without turning laterally out to the IS nerve, even though the IS nerve is present. For a summary of the aCC mutant phenotypes, see Fig. 2.

Methods. To allow normal *eve* stripe expression while selectively inactivating *eve* CNS function, we used the temperature-sensitive *eve*^{1D19} allele²⁸. To specifically inactivate *eve* CNS function, embryos homozygous for *eve*^{1D19} were raised at 18°C through the blastoderm stage, shifted to 30°C between 2.5–4.0 h of development, and raised at 30°C throughout neurogenesis.

each segment is not formed²⁸. The CNS is consequently greatly reduced and aberrant. Thus, normal *eve* stripe expression is a prerequisite for analysing *eve* CNS function. To allow blastoderm stripe expression while selectively inactivating *eve* CNS function, we used the temperature-sensitive *eve*^{1D19} allele²⁸. At the permissive temperature (18°C), embryos homozygous for *eve*^{1D19} (subsequently called *eve*^{ts} embryos) are not viable but they do show normal epidermal and neuronal development in all segments but the second and sixth abdominal (A2 and A6), where part of each segment is lacking. At the non-permissive temperature (30°C), *eve*^{ts} embryos show a phenotype identical to null *eve* alleles²⁸, indicating that at 30°C the *eve*^{ts} protein is substantially or completely inactivated.

By raising *eve*^{ts} embryos at 18°C through the blastoderm stage, and then shifting the temperature to 30°C throughout neurogenesis, we can specifically inactivate (either partly or completely) the *eve* CNS function. Embryos shifted to 30°C after 2.5 h of development have an identical pattern of segmentation to embryos raised continuously at 18°C, demonstrating that blastoderm function is not affected by a temperature shift at this time. In all the experiments described below, we removed *eve* CNS function by raising a collection of *eve*^{ts} embryos at 18°C until they were 2.5–4.0 h old, and then shifted them to 30°C until they were old enough to assay the extent of CNS development.

To determine whether neurons normally expressing *eve* protein (called *eve*⁺ neurons) survive in the absence of *eve* CNS function, and to determine their cell body position, we stained *eve*^{ts} embryos lacking *eve* CNS function with an antiserum against the *eve* gene product²¹. The *eve*^{ts} embryos were identified either by defects in A2 and A6 segments, or by the lack of β -galactosidase expression from a *ftz* promoter-*lacZ* gene construct¹⁵ on the balancer chromosome (only homozygous *eve*^{ts} embryos lack the *ftz-lacZ* balancer chromosome, and hence these embryos are unstained by an antibody to β -galactosidase). In embryos lacking *eve* CNS expression, *eve*⁺ neurons accumulate *eve* protein which is presumably inactive (data not shown). In these embryos, the *eve*⁺ neurons can be identified in their characteristic positions, indicating that the *eve*⁺ neurons survive

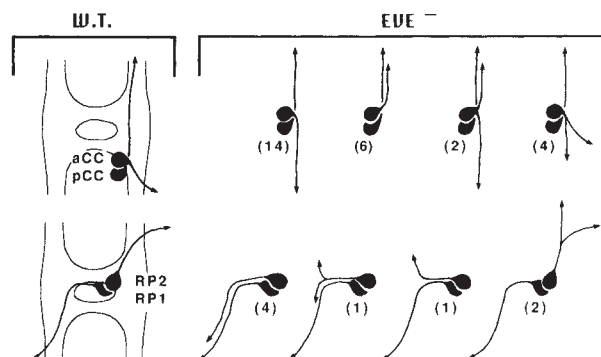


Fig. 2 CNS expression of *eve* is required for the development of the normal axon morphology of the aCC and RP2 neurons. Drawings of Lucifer Yellow injections of identified *eve*⁺ neurons in wild-type embryos and in embryos lacking *eve* CNS expression. The number of cases observed of each wild-type axon morphology is >40. The number of cases of a particular axon morphology in embryos lacking *eve* CNS expression is shown in parentheses beneath each result. The neuropil is outlined in the wild-type panel. Top row shows that of the aCC and pCC neurons, only the aCC neuron shows an aberrant axon morphology in the absence of *eve* CNS expression. The bottom row shows the RP1 and RP2 neurons; only the RP2 neuron shows an aberrant axon morphology in the absence of *eve* CNS expression.

in the absence (or reduction) of functional *eve* protein and that the cell body positions of these neurons are unaltered. Thus we can unequivocally identify *eve*⁺ neurons in *eve*^{ts} embryos.

To assay the role of *eve* during neuronal development, we initially stained embryos lacking *eve* CNS function with a serum antibody which recognizes all axons²⁹. No difference in the overall organization of the longitudinal and commissural axon pathways in the CNS or peripheral nerve roots was observed. Thus, we subsequently assayed the morphology of the *eve*⁺ neurons within the context of an otherwise normal CNS. We assayed the fate of three *eve*⁺ neurons (aCC, pCC, and RP2), and an *eve*⁻ sibling (RP1), by Lucifer Yellow injections²⁴ to reveal their morphology.

The sibling aCC and pCC neurons are born from NB 1-1 in the anterior part of each segment, and migrate anteriorly to lie just posterior to the posterior commissure in the adjacent segment^{3,9}. The aCC is a motoneuron which normally extends its axon laterally out into the periphery, pioneering the intersegmental (IS) nerve^{24,25}, and ultimately synapsing with a body wall muscle 1³⁰. The pCC is an interneuron which extends an axon anteriorly in one of the longitudinal axon fascicles. The RP2 axon normally extends anteriorly and laterally out of the ipsilateral IS nerve of the adjacent anterior segment, whereas its sibling RP1 extends its axon across the midline and then posteriorly and laterally out of the contralateral IS nerve of its same segment^{25,26}. The axon morphology of these four neurons is unaltered in *eve*^{ts} embryos grown continually at 18°C.

In embryos lacking *eve* CNS function, the normally *eve*⁺ pCC and the *eve*⁻ RP1 continue to display wild-type axonal morphology. In contrast, the normally *eve*⁺ aCC and RP2 neurons show aberrant axonal morphology (Figs 1, 2). Although the aCC always migrates to its appropriate position, its axonal morphology is never normal. The most frequent aCC phenotype (14/26 cases) is to extend an axon posteriorly without turning laterally out of the IS nerve (Figs 1, 2). Other aCC neurons show a variety of phenotypes, including extending an axon anteriorly (6/26 cases), extending axons both anteriorly and posteriorly (2/26 cases), and extending short axons both posteriorly and laterally towards the IS nerve (4/26 cases). Thus,

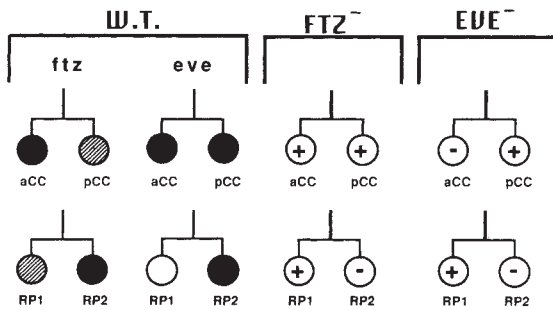


Fig. 3 CNS expression of *eve* and *ftz* is required for the correct determination of identified neurons. Summary showing the overlapping expression of two segmentation genes, *ftz* and *eve*, in four identified neurons (aCC, pCC, RP1, and RP2), and the effects of eliminating the CNS expression of each of these genes on the axon morphology of these four neurons. +, normal axon morphology; - abnormal axon morphology. These results (here and ref. 3) suggest that the nuclear protein products of the *eve* and *ftz* segmentation genes are components of the mechanism controlling cell fate during neuronal development in *Drosophila*.

in 22/26 aCC neurons assayed in mutant embryos, the aCC growth cone did not turn laterally to pioneer the IS nerve, although the IS nerve formed in all of these cases. These results suggest that in the absence of *eve* CNS function, the cues that establish the IS nerve are still present, but the aCC growth cone cannot properly recognize them. In the absence of the aCC axon, other neurons pioneer the IS nerve; this is not surprising, as it occurs in other species^{25,30}.

Following removal of *eve* CNS function, we also observed transformation of the RP2 neuron. The most frequent RP2 phenotype is to extend its axon contralaterally, paralleling the pathway of its sibling RP1 axon (Fig. 2). Thus, loss of *eve* function apparently transforms both the aCC and RP2 neurons, each showing a preferred alternate morphology as well as a number of other, less frequent phenotypes.

The aberrant development of the aCC and RP2 neurons is almost certainly due to the loss of *eve* function in these cells, rather than their response to an altered environment. Expression of *eve* cannot be detected in or around the CNS except in 16 neurons per hemisegment (it is also expressed in cells at the dorsal surface of the embryo and in the anal pad)²¹. Defects in other neurons should not affect the aCC or RP2 morphology, as they are among the first neurons to extend axons out of the CNS^{24,25}. The non-neuronal guidance cues used by the aCC and RP2 growth cones appear unaltered, as nerves form in the correct positions despite loss of the aCC and RP2 axons. Thus, loss of *eve* function in the aCC and RP2 neurons appears to produce specific transformations of neuronal fate.

In addition to assaying axon morphology, we used antibody probes to assay the CNS expression of several other segmentation genes in embryos lacking *eve* CNS expression. The segmentally repeated expression of the *ftz* protein in the aCC, pCC, RP1, and RP2 neurons³ is unaltered despite loss of *eve* expression. Thus, regulatory interactions between *eve* and *ftz* in the CNS are different from those in the blastoderm; loss of *eve* function at the blastoderm stage clearly alters the striped pattern of *ftz* protein distribution²⁷. Different blastoderm and CNS regulatory interactions are also observed between *eve* and the segmentation gene *engrailed* (*en*). Loss of *eve* blastoderm function causes loss of most epidermal *en* expression²⁰; loss of *eve* CNS function, however, has no effect on the CNS protein pattern of *en* (data not shown).

The three identified *eve*⁺ neurons described in this paper (aCC, pCC, and RP2) are also *ftz*⁺ neurons³ (Fig. 3). All three neurons have different responses to the loss of *ftz* and/or *eve*

CNS function (Fig. 3). The pCC neuron is refractory to loss of function of either gene, suggesting that other genes are involved in regulating the axon morphology of this neuron. In the absence of *ftz* expression, the sibling aCC neuron also shows no change in either *eve* expression or axon morphology³; the lack of *eve* expression, however, does result in aberrant aCC axon morphology (Figs 1, 2). The third neuron, RP2, shows the most interesting responses. In the absence of *ftz* CNS expression, RP2 does not express *eve* and shows an aberrant axon morphology (mimicking RP1 morphology)³; in the absence of *eve* CNS expression, the RP2 neuron continues to express *ftz* but develops a similar RP1-like aberrant axon morphology. Thus loss of *eve* alone leads to the same RP2 phenotype as loss of both *ftz* and *eve*. This suggests a hierarchical interaction in the RP2 neuron, with *ftz* activating *eve*, and then *eve* regulating genes which control the axonal morphology of RP2.

The interaction of *ftz* and *eve* with upstream regulatory genes, and with the downstream effector genes responsible for neuronal differentiation, are likely to be complex and possibly unique to individual neurons. Candidate downstream genes have been identified (such as fasciclin III, ref 26) and their interactions with *ftz* and *eve* are now testable. In addition to *ftz* and *eve*, many other genes controlling the pattern of *Drosophila* segmentation are subsequently expressed during neurogenesis, as are a growing number of vertebrate homologues to these genes³¹⁻³⁷. We think it is likely that other segmentation genes are also involved in the generation of neuronal diversity in *Drosophila*, and that their homologues play similar roles in vertebrates.

We thank Yash Hiromi and Nipam Patel for the β -gal balancer chromosome and *eve* β -gal stock, Manfred Frasch and Michael Levine for anti-*eve* antibodies, Henry Krause and Walter Gehring for anti-*ftz*, and Kevin Coleman and Tom Kornberg for collaboration in making anti-*en*. Supported by NICHD Developmental Biology Traineeship to C.Q.D., NIH Postdoctoral Fellowship to D.S., and NIMH and NIH grants to C.S.G.

Received 15 February; accepted 14 April 1988.

- Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
- Lewis, E. B. *Nature* **276**, 565-570 (1978).
- Doe, C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. *Science* **239**, 170-175 (1988).
- Bate, C. M. *J. Embryol. exp. Morph.* **35**, 107-123 (1976).
- Doe, C. Q. & Goodman, C. S. *Dev. Biol.* **111**, 193-205 (1985).
- Doe, C. Q. & Goodman, C. S. *Dev. Biol.* **111**, 206-219 (1985).
- Hartenstein, V. & Campos-Ortega, J. A. *Wilhelm Roux Arch. dev. Biol.* **193**, 308-325 (1984).
- Goodman, C. S. & Spitzer, N. C. *Nature* **280**, 208-214 (1979).
- Goodman, C. S., Raper, J. A., Ho, R. K. & Chang, S. in *Developmental Order and Regulation* (eds Subtelny, S. & Green, P. B.) 275-316 (Liss, New York, 1982).
- Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* **3**, 20-30 (1983).
- Taghert, P. H. & Goodman, C. S. *J. Neurosci.* **4**, 989-1000 (1984).
- Kuwada, J. Y. & Goodman, C. S. *Dev. Biol.* **110**, 114-126 (1985).
- Lehmann, R., Jimenez, F., Dietrich, U. & Campos-Ortega, J. A. *Wilhelm Roux Arch. dev. Biol.* **192**, 62-74 (1983).
- Campos-Ortega, J. A. *Trends Neurosci.* **8**, 245-250 (1985).
- Hiromi, Y., Kuroiwa, A. & Gehring, W. J. *Cell* **43**, 603-613 (1985).
- Carroll, S. B. & Scott, M. P. *Cell* **43**, 47-57 (1985).
- DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59-69 (1985).
- Knipple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A. & Jackle, H. *Nature* **317**, 40-44 (1985).
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., Noll, M. *Cell* **47**, 1033-1040 (1986).
- MacDonald, P. M., Ingham, P. & Struhl, G. *Cell* **47**, 721-734 (1986).
- Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. *EMBO J.* **6**, 749-759 (1987).
- Tautz, D. *et al.* *Nature* **327**, 383-389 (1987).
- Baker, N. E. *EMBO J.* **6**, 1765-1773 (1987).
- Thomas, J. B., Bastiani, M. J., Bate, M. & Goodman, C. S. *Nature* **310**, 203-207 (1984).
- Goodman, C. S. *et al.* *Science* **225**, 1271-1279 (1984).
- Patel, N. H., Snow, P. M. & Goodman, C. S. *Cell* **48**, 975-988 (1987).
- Carroll, S. B. & Scott, M. P. *Cell* **45**, 113-126 (1986).
- Nusslein-Volhard, C., Kluding, H. & Jurgens, G. *Cold Spring Harb. Symp. quant. Biol.* **50**, 145-154 (1985).
- Jan, L. Y. & Jan, Y. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2700-2704 (1982).
- Bastiani, M. J., Doe, C. Q., Helfand, S. L. & Goodman, C. S. *Trends Neurosci.* **8**, 257-266 (1985).
- Awgulewitsch, A., Utset, M. F., Hart, C. P., McGinnis, W. & Ruddle, F. H. *Nature* **320**, 328-335 (1986).
- Simeone, A. *et al.* *Nature* **320**, 763-765 (1986).
- Carrasco, A. E. & Malacinski, G. M. *Dev. Biol.* **121**, 69-81 (1987).
- Joyner, A. L. & Martin, G. R. *Genes Dev.* **1**, 29-38 (1987).
- Odenwald, W. F. *et al.* *Genes Dev.* **1**, 482-496 (1987).
- Krumlauf, R., Holland, P. W. H., McVey, J. H. & Hogan, B. L. M. *Development* **99**, 603-617 (1987).
- Fainsod, A., Awgulewitsch, A. & Ruddle, F. H. *Dev. Biol.* **124**, 125-133 (1987).

Computer-based tutorials take off

S. Smith

Computers have come to the science classroom with the increasing popularity of tutorial programs, video simulations and interactive video discs.

ALTHOUGH computer-assisted instruction (CAI) was developed using software run on main-frame computers¹, most current applications employ microcomputers to take advantage of their speed, low cost, and ease of use. Tutorial programs which run on microcomputers may be used by students on individual machines² or on machines which are networked, to provide automatic record-keeping of student use and performance³.

Microcomputers also may be used for data collection and analysis, both in the laboratory⁴ and in the classroom, by connecting a microcomputer to a video projector so that large numbers of students may view the output. For example, Meiss⁵ has developed a program on the behaviour of active skeletal muscle where the instructor can control the speed of the presentation and the preload muscle length, afterload muscle length, and contractility. The computer plots simultaneously the relationships among force, length, and time, with primary focus on the length-tension diagram.

Programs which are designed to help students learn a specific topic have many forms and objectives. Attempts to classify lessons as tutorial, simulation or drill-and-practice are generally over-simplifications of the actual instructional strategies employed, which are often a complex of all three approaches. For example, a very simple lesson to help students learn the names of the elements is facilitated by a "drill" on the element names and corresponding symbols. Although simple in concept, the effectiveness of such programs is enhanced by incorporating into the drill program an automatic review of the names the student got wrong. Reviews could include specific feedback on spelling errors, such as missing, extra, and wrong or inverted letters, and comment on the correct names for the mistaken element, such as pointing out why the symbol "I" stands for iodine instead of iron. Programs which provide extensive feedback and tutoring based on student input are sometimes called intelligent computer-assisted instruction (ICAI), or intelligent tutors⁶.

Graphic demonstration

One of the advantages of computer-assisted instruction is that text can be augmented by computer graphics to allow students to visualize complex systems.



Fig. 1 An example of an interactive video disc tutorial. Students touch the touch-sensitive video display to indicate their response to the question.

Tutorials coupled with graphics help students to understand what is happening by letting them see immediately the consequences of changing the parameters which define a system. Programs developed by Schatz⁶ allow students to simulate the operation of NMR and IR instruments by drawing the actual spectra on the computer screen which result from the machine settings specified by the student. Specific help is provided where needed.

Interactive video discs

Actual television images can be integrated into traditional computer-based instructional programs by the use of a video disc player controlled by the computer⁷. A video disc player provides random access to 54,000 still frames, yielding up to 30 minutes of full-motion video on one side of a 12-inch video disc. The images can be displayed on a separate screen from the computer-generated material or be mixed with computer graphics on one screen⁸. One-screen systems also can be configured to allow video images to be displayed within a window in the computer-generated material. Having video and computer material on the same screen combines the advantages of real television images with the interactive qualities of traditional CAI to produce a new instructional medium. The use of television pictures in interactive video disc lessons provides consider-

able realism to instructional programs and allows students to do experiments which are either too expensive, complex, hazardous, or time-consuming to be performed in the actual laboratory⁹. Random access to video images as either still frames or motion sequences allows students to try various operations and immediately to see the result. This flexibility is likely to increase the use of interactive video instruction using the analogue images of a video disc player or digitized images¹⁰. □

Stanley Smith is at the Department of Chemistry, University of Illinois, 1209 West California Street, Urbana, Illinois 61801, USA. For more information, fill in reader service number 100.

1. Smith, S.G. & Sherwood, B.A. *Science* **192**, 344-352 (1976).
2. Moore, J., et al. *J. chem. Educ.* **56**, 776-783 (1979).
3. Jones, L., Karloski, J. & Smith, S. *Academic Computing* **2**, 36-37 (1987).
4. Ward, C.R. & Reeves, J.H. *MicroComputer Applications in Chemistry* (Scott Foresman Little Brown, Glenview, Illinois, 1988).
5. Meiss, R.A., *Cardiac Muscle Mechanics* (Compress, Wentworth, New Hampshire, 1987).
6. Anderson, J.R., Boyle, C.F. & Reiser, B.J. *Science* **228**, 456-462 (1985).
7. Schatz, P.F. *Academic Computing* **2**, 38-39 (1987).
8. Parsole, E. *Interactive Video*, (Sigma Technical Press, Wilmslow, Cheshire, U.K., 1983).
9. Brooks, D.W., Lyons, E.J. & Tipson, T.J. *J. chem. Educ.* **62**, 514-515 (1985).
10. Smith, S.G., Jones, L.L. & Waugh, M.L. *J. Computer-based Instr.* **4**, 117-121 (1986).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Tools for teaching science

Video disc programmes, slide presentations and teaching microscopes are all part of this week's review of science teaching aids.

THE Cy-Scan cytology training system from E. Leitz Instruments Ltd was designed to help histopathology and cytology laboratories cope with training new personnel to accommodate the



increased workload of cervical screening (Reader Service No. 101). The £30,000 (UK) Cy-Scan laboratory training system consists of a computer connected to an X-Y coordinate readout unit fitted to a normal Leitz microscope. The integrated software and hardware allow the trainee to scan slides previously checked by a qualified screener and have their results automatically checked against those of the qualified screener or pathologist. It can also be used as a quality control system to check examples from the routine daily work of the laboratory.

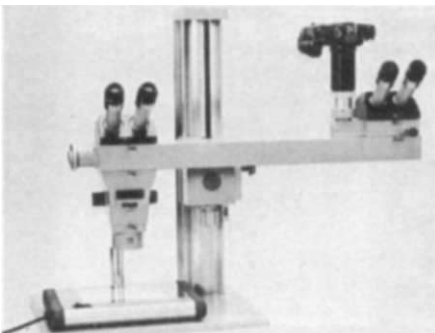
ISCO, Inc. offers a video disc programme entitled *Biotechnology: Learning the Techniques* that provides step-by-step instruction on 33 protocols commonly used in the biotechnology laboratory (Reader Service No. 102). The \$995 (US) self-paced **biotechnology video disc tutorial** includes a detailed manual. It shows beginners how to do procedures right the first time, and provides the means for experienced workers to improve and standardize their techniques. The laser video disc format provides instant location and replay of any lesson. Programmes can



A video disc for learning biotechniques.

be viewed clearly at regular speed or in slow motion, and can be frozen for a closer view. Among the topics included are: ligation, plasmid isolation, restriction analysis, mouse immunization, ELISA, purification of monoclonal antibodies, affinity chromatography, SDS polyacrylamide gel electrophoresis, and Western blots.

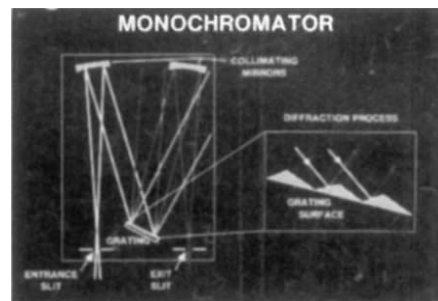
Zeiss has a new dual-observation **teaching stereomicroscope** with continuously variable 1:8 zoom magnification (Reader Service No. 103). Zeiss says the \$15,350 (US) model SV8 is ideal for training and instruction in colleges and workshops, exacting repair work, and the observation and photodocumentation of microscopic specimens. The Zeiss microscope offers identical observation conditions and a



Two heads are better than one with the Zeiss microscope.

25 mm field of view to both observers, who sit side-by-side, spaced 60 cm apart. Stability is ensured by the stand assembly, which has a 640 mm-high column with counterweight. The microscope has a photographable light pointer which can be rotated 360° and moved over the entire field of view. Three objectives are provided: a standard 100 mm objective, a DS 100 mm dual objective that converts the microscope to a Greenough-type stereomicroscope, and a flatfield 100 mm objective for viewing low-profile objects.

Savant's extensive line of audiovisual training packages has been expanded to include a **programme on UV-visible spectroscopy** (Reader Service No. 104). The 36-minute programme, entitled *An Introduction to Ultraviolet-Visible Spectroscopy*, comes in a slide presentation format for \$265 (US), and a video format for \$540 (US). The basics discussed in the introductory programme include: fundamental properties and characteristics of electromagnetic radiation; wavelength, frequency and energy relationships; the absorption process; energy-level dia-



Savant's tutorial on UV-visible spectroscopy. grams; chromophores; conjugation; auxochromes; spectral shifts; and pH and solvent effects. In the quantitative section, the principles of Beer's Law are developed, and an actual quantitative experiment illustrates how an unknown concentration is determined. Instrumentation is reviewed in a generic fashion, covering sources, monochromators, detectors and readout devices.

GOW-MAC Instrument Company offers a 5-hour **audio course on gas chromatography** entitled *Basic Gas Chromatography* (Reader Service No. 105). A 218-page course manual is included with the six audio cassettes, and the entire package costs \$400 (US). The course covers the basic principles required to understand and improve GC separations, methods of selecting proper equipment, operating parameters for good separation results, and the latest developments in columns, instrumentation and analysis.

Science Media, a division of J. Huley Associates, Inc., is offering a set of three slide-tape presentations on the **interpretation of NMR spectra** (Reader Service No. 106). Two programmes are devoted to proton absorbance, covering simple splitting to more complex spin-spin splitting patterns. The third programme describes ¹³C chemical shifts, the shielding constant and factors influencing its magnitude, and the subject of assignment techniques. The educational and training programmes were developed as a reference and quick refresher for the professional chemist, a learning tool for the student, and a means to broaden the skills of the laboratory technician. The \$150 (US) slide presentation course is also available on videotape for \$295 (US). □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.